Columbus State University
CSU ePress

Theses and Dissertations

5-2017

The Impact of Teacher-Student Relationships and Classroom Engagement On Student Growth Percentiles of 7th and 8th Grade Students in One Rural School in Southwest Georgia

Dave Dennie

Follow this and additional works at: http://csuepress.columbusstate.edu/theses_dissertations

Part of the Educational Leadership Commons

Recommended Citation
Dennie, Dave, "The Impact of Teacher-Student Relationships and Classroom Engagement On Student Growth Percentiles of 7th and 8th Grade Students in One Rural School in Southwest Georgia" (2017). *Theses and Dissertations*. 223.
http://csuepress.columbusstate.edu/theses_dissertations/223

This Dissertation is brought to you for free and open access by CSU ePress. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of CSU ePress.

LETTER OF TRANSMITTAL

The Letter of Transmittal is the first page of the dissertation, but should not be numbered. You must submit a Letter of Transmittal for each of the 8 required copies of the dissertation. Each signature on each page should be original in blue ink. The Letter of Transmittal page will appear exactly as shown in the sample attached to this document. Since the Letter of Transmittal page represents an administrative action and is not part of the dissertation, it is not listed in the table of contents. However, it is bound with the dissertation

THE IMPACT OF TEACHER-STUDENT RELATIONSHIPS AND CLASSROOM ENGAGEMENT ON STUDENT GROWTH PERCENTILES OF 7TH AND 8TH GRADE STUDENTS IN ONE RURAL SCHOOL IN SOUTHWEST GEORGIA

By
Dave Dennie

A Dissertation
Submitted to the Faculty of
Columbus State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Education
in Curriculum and Leadership
(Curriculum)

Columbus State University
Columbus, GA

May 2017

DEDICATION

I dedicate this dissertation to my parents Karen and Dave. You have supported me through the good and bad in life and expected only the best. Your patience and perseverance through my many missteps in life has shaped me and allowed me to be who I am. You have taught me to work hard for the things that I aspire to achieve. Thank you for always being there for me.

I also dedicate this dissertation to my coach, my boss, my friend, Tim Habecker. You taught me to be a man and take responsibilities for my actions and encouraged me to go to college. As I reflect back on my life, I do not know if I am where I am without you in my corner. You changed the course of my life and words cannot describe how thankful I am.

Finally, I dedicate my dissertation to my wife, Kim, who has been by my side providing support and encouragement during my many years of schooling. I am grateful for your patience with the amount of time I had to put in to complete my dissertation, which made the process less stressful. I am lucky to have you in my life.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Wyndol Furman for granting permission to use the Network of Relationships Inventory, Dr. Jennifer La Guardia for granting permission to use the Basic Psychological Needs Inventory, Dr. Ze Wange for granting permission to use the Classroom Engagement Inventory, and Dr. James Connell for granting permission to reprint the Self-Systems Process Model.

I would like to acknowledge Dr. James Martin and Mrs. Stacey Carlisle for granting permission to conduct this study at Harris County Carver Middle School. A special thanks to Mr. Carl Dekker for helping with parental consent forms and scheduling time with students.

I would like to acknowledge Dr. Christy Cabezas for encouraging me to go back to school and continue my education towards a higher degree. Your words of wisdom started me on this path.

I would also like to acknowledge my editor, Dr. Donna Patterson, who spent many hours proofreading and providing constructive feedback on my chapters. Your experience, knowledge, and flawless grammatical editing were invaluable.

ABSTRACT

A number of states throughout the United States, including Georgia, are implementing a relatively new metric, student growth percentiles, as part of teacher and leader evaluations. Student growth plays a tremendous role in evaluations, accounting for up to 50% of a teacher or leader evaluations, yet there is little to no peer reviewed research on classroom factors that influence student growth percentiles.

This quantitative study examined the extent that teacher-student relationships influenced basic psychological needs, engagement, and student growth using the Self-systems Process Model as a framework using structural equation modeling. Based on prior research, it was hypothesized that context (teacher-student relationship) influenced self (basic psychological needs), which influenced action (engagement), and consequently, influenced outcome (outcome).

At the end of the 2015-2016 school year, data was collected from seventh and eighth grade students in a medium to large school district in southwest Georgia that was 73.4% white, 16.6% African-American, 3.4% Hispanic, and 5.1% multiracial, with 29.7% of the students receiving free lunch and 6.3% of the students receiving reduced-price lunch. The 512 student responses were representative of the school population.

Student responses to the modified Network of Relationships Inventory, Basic Psychological Needs Inventory, and Classroom Engagement Inventory showed that students perceived the following: that there was a positive teacher-student relationship, that their basic psychological needs were satisfied in the classroom, and that they were

actively engaged. Student responses and their outcomes of Georgia Milestones standardized assessment norm-referenced scores, scale scores, student growth percentiles, and class GPA were used to complete a structural equation modeling analysis.

The findings of the study supported prior research that a positive teacher-student relationship positively influenced levels of engagement in the classroom and, consequently, student outcomes as measured by classroom GPA and standardized assessment results. Using an identical methodological setup that substituted student growth percentiles for scale scores, it was determined that teacher-student relationships, basic psychological need satisfaction, and level of engagement do not influence student growth percentiles across socioeconomic levels and race.

TABLE OF CONTENTS

Table of Contents (continued)

Table of Contents (continued)

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

Accountability Changes

Social Efficiency advocates hold assessment and evaluation in high regard (Schiro, 2013). Evaluations can and are used to assess students, teachers, schools, and the curriculum. Assessments provide student feedback toward meeting standards, teacher feedback of effectiveness or lack thereof of instruction, and feedback to school administration on the teaching and learning process (Joshua, Joshua, & Kritsonis, 2006). Student scores are used to rate students, teachers, principals, and schools in terms of performance and effectiveness because everyone in the process is accountable for the terminal objectives. Educators are the clients of the public and, therefore, are accountable to the public (Schiro, 2013).

Prior to 2000, states developed their own assessment systems to determine and track students' levels of achievement (Schafer, Lissitz, Zhu, & Zhang, 2012). Starting in 2001, aligning with the Social Efficiency ideology of educating students for the public in order to better serve society (Schiro, 2013), with the implementation of the No Child Left Behind Act (NCLB), states were required to measure school status (levels of achievement) every year (Schafer et al., 2012). The ultimate goal of NCLB was to ensure proficiency of all students by 2016 in reading and math, which was identified through the use of student achievement measures on standardized assessments (Nichols, Glass & Berliner, 2005; Ladd & Lauen, 2010).

With a Social Efficiency mindset along with the NCLB legislation, use of status models assisted educators, legislators, and the public to identify the percentage of students who met or did not meet the stated objectives (Thurlow, Lazarus, Quenemoen & Moen, 2010). Status models were a simple representation of student achievement levels based upon the state's predefined performance standards, and they required an acceptable level of achievement from all students, regardless of prior academic achievement (Betebenner, 2008).

Teaching within the Social Efficiency Ideology (SEI) required teachers to be orchestrators of the classroom, pushing students in the right direction, encouraging them, evaluating them, and providing them prompt feedback (Schiro, 2013) in order to get them to grow as students (Thurlow et al., 2010). Teachers are the most important contributor to student learning within the school, typically accounting for 9% to 13% of the variance in student achievement (Haertal, 2013). Wentzel (2002) went as far as stating that teachers may influence students' motivation and behavior more than parents. Ultimately, teaching under NCLB, a standards-based accountability system, was evaluated by percentage of students meeting the standards with the goal to improve student achievement based on standards (Ladd & Lauen, 2010). Standards-based curriculum calls for improved outcomes, and standardized test scores should improve if educational quality improves (Doran, 2003). If scores do not go up, educational quality does not improve, and teachers and schools should be held accountable (Haertel, 2013).

Ladd and Lauen (2010) stated, "With standards-based accountability programs, policymakers set clear standards, measure student performance, and use those measures to evaluate the effectiveness of schools" (p. 426). They further explained, "The theory of

action behind educational accountability is that setting standards and measuring performance relative to standards will lead to teachers working harder and students learning more” (p. 427) to which Schiro (2013) also alluded in his research. Assessment results should provide information on how to improve educational results, but statistical methods on NCLB have failed to do so (Doran, 2003).

Using achievement scores, which are snapshots of student ability at the end of the year, to rate teachers is inappropriate. Some students come to class missing prerequisite skills needed to be successful, which is counter to the SEI (Schiro, 2013). While the reason students are unprepared are numerous, the results are the same. Students are moving through school, and upon graduation, are not prepared to be successful functioning members of society, which is driving current educational reform (Schiro, 2013).

One of the major issues of using achievement scores to evaluate schools, administrators, and teachers is that achievement scores include both school and non-school effects, which are out of a school's control (Joshua et al., 2006; Ladd & Lauen, 2010). How can a teacher be held accountable for a student's achievement when the student came in lacking prerequisite skills needed to be successful in class? If teachers truly account for only 9% to 13% of a student's achievement (Haertal, 2013) within a school year, and a student is lacking 25% of the knowledge needed to be successful, why should the teacher be penalized if the child grew while in the teacher's charge? In the SEI framework, someone must be accountable for failure to meet standards (Schiro, 2013), and since the teacher is the most important determinant in student growth within the classroom, the teacher should be held accountable (Haertel, 2013) for growth while the

student is under his or her charge. This rationale was the framework from which Race to the Top (RttT) was born.

In the past 10 years due to RttT and the Growth Model Pilot Program (GMPP), accountability has shifted from a focus on effectiveness of schools to a focus on effectiveness of teachers within the school. This shift was accompanied by the use of growth models along with status scores (Collins & Amrein-Beardsley, 2014; USDOE, 2009). Prior to RttT, state teacher evaluation systems lacked rigor, with many evaluation systems placing all teachers at or near the top (Goldhaber, Walch, & Gabele, 2014). Failure to recognize the differences among teachers in an evaluation system creates a situation in which decision making is difficult, as there is no variation in performance. Secretary of Education Arne Duncan pointed this out when he stated, “Today in our country, 99% of our teachers are above average,” (Gabriel, 2010, September 2) indicating there were obvious problems with teacher evaluation systems. Under the RttT requirement, of which Georgia is a part, states were required to improve teacher and leader effectiveness by developing a robust evaluation system, which included a variety of sources of information to inform of teacher and leader effectiveness.

Requirements of RttT included the following: A way to measure the growth of every student individually, an evaluation system that takes into account multiple measures of teacher effectiveness such as administrator ratings, student growth, and student surveys, with student growth being a significant factor of evaluation, annual evaluations and feedback on student growth, and a plan to use evaluations and growth data to inform decisions regarding professional development, certification, compensation, and other various incentives and sanctions (USDOE, 2009).

Under the RttT implementation, school districts across the state of Georgia were required to change educator evaluations to include student growth and multiple classroom observations (O.C.G.A. § 20-2-210(b), 2013) under the system known as Teacher Keys Effectiveness System and Leader Keys Effectiveness System (TKES and LKES). Teachers were no longer solely evaluated by the school administrator two or three times a year, but six times a year using a rigorous administrator evaluation along with a student growth metric (GaDEO-OSI, 2014a). Prior to the implementation of TKES, levels of student achievement were measured and used only as accountability data at the school and district level, not at the administrator or teacher level. While administrators may have looked at classroom achievement results by teacher, there was no mention of student achievement or accountability based on student achievement in the teacher evaluation process.

TKES is a multidimensional look at teacher effectiveness and includes the two components of rigorous administrator evaluation and aggregated student growth scores, with each being rated from level I to level IV (GaDEO-OSI, 2014a). According to the *Teacher Keys Effectiveness System* manual, the administrator portion, referred to as Teacher Assessment on Performance Standards (TAPS) is determined by evaluators using a qualitative rubrics-based evaluation tool based on ten performance standards, through six classroom observations throughout the school year. According to the results of all six walkthroughs, a teacher is assigned an overall TAPS rating of I (Ineffective), II (Needs Development), III (Proficient), or IV (Exemplary) as seen on the horizontal axis of the matrix in Figure 1.

The student growth portion, referred to as Overall Student Growth Rating, is determined by aggregating student growth scores as determined by using student growth percentiles based on state standardized assessment results. The process will be described further in depth in the review. The teacher is assigned an overall student growth rating of I, II, III, or IV as seen on the vertical axis of the matrix in Figure 1.

Overall Student Growth Rating				
Level IV	Needs Development	Proficient	Exemplary	Exemplary
Level III	Needs Development	Proficient	Proficient	Exemplary
Level II	Ineffective	Needs Development	Needs Development	Proficient
Level I	Ineffective	Ineffective	Needs Development	Needs Development
	Level I	Level II	Level III	Level IV

Overall TAPs Rating

Figure 1. Teacher effectiveness measure (TEM) matrix. Reprinted from "TEM scoring guide and methodology." GaDOE, 2014b, p. 20. Retrieved from http://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Documents/TEM%20Scoring%20Guide%20206-18-14Final.pdf

According to the 2014 TKES handbook, in the Georgia model, student growth ratings account for fifty percent of the TEM with the other fifty percent being accounted for by overall TAPS rating. Clearly, student growth plays a tremendous role in teacher evaluations, bringing to the forefront, the need to evaluate factors that influence student growth in the classroom.

Teacher expectations are that students progress and grow towards achieving terminal objectives. Currently, while growth scores represent a significant portion of a teacher's evaluation in Georgia, there is little research on variables that influence student growth. Logically, it can be assumed factors that influence student achievement also

influence student growth. If gain scores were used to determine student growth from beginning to end, that may be the case; however, using student growth percentiles presents a problem because the logic behind achievement/status and growth as determined by student growth percentiles (SGPs) are quite different. Visual inspection of status scores are easily interpretable as they are high or low, passing or not, as long as the interpreter of the scores knows the cut scores and maximum and minimum scores as these values are absolute. Correlating achievement to interventions was easily determined, as interpreters of the scores would see scores go up or down. If an intervention worked positively, scores would go up; if not, scores would stay flat or go down.

While the interpretation of SGPs is not difficult, how student growth is impacted by interventions is more complex and not as logically interpreted. It is possible that two students have a growth score of 60th percentile, yet result in totally different interpretation as these scores are relative. Student growth using SGPs are determined based on at least two years of test scores and comparisons to similar achieving students. A student that achieved a status score of 750 last year on the math CRCT will be compared to all students in the state of Georgia who also scored a 750 (GaDOE-CIA, 2014b). All these students attend various schools throughout the state with various cultures, school climate, home structures, support levels, and teachers. On this year's standardized assessment, the student scores an 850 and is still compared to his/her similar academic peers from last year. Because the assessments are not vertically aligned, the 100 point gain is disregarded since there is no meaning in a norm-referenced system. If all other students being compared to this student scored an 860, this student will have lower growth. If the academic peers scored lower, this student will have higher growth.

How did some intervention affect growth, if part of the reason growth occurred depends on other students throughout the state? To further illustrate the problem, it may be possible that the student in question had high levels of engagement, which has been shown to improve achievement, yet the other students had higher levels of engagement, thereby having higher levels of achievement, making our student's growth score lower. A factor that influences achievement may not influence growth as determined by the state of Georgia, which can have a major impact on a teacher's evaluation.

While student achievement is still important in the evaluation of schools and teachers, growth scores are now at the forefront of concern of teachers in at least forty states (Collins & Amrein-Beardsley, 2014) as the stakes have never been higher for teachers with evaluations being tied to student growth. Many states now reward and penalize teachers based on the amount of their students' growth (Barnett & Amrein-Beardsley, 2011; Schochet & Chiang, 2010). While Georgia has not specified the repercussions for teachers who are consistently rated at Level II or lower, ideas have surfaced that consist of putting teachers on professional development plans, losing state certification, not receiving performance pay, and even termination for teachers who chronically achieve low teacher effectiveness measures (TEM).

Self-Determination Theory

With a portion of teacher and leader accountability now based on student growth as determined by student growth percentiles, research of factors, both proximal and distal, that may impact student growth is needed, as research of factors that influence student achievement may not be applicable. Self-determination theory (SDT) is a well-supported theory of motivation and engagement and has been shown to influence student

achievement. SDT is a hypothesized model developed to understand and explain intrinsic motivation in individuals (Deci & Ryan, 1985). Deci and Ryan posited that all individuals are active and growth oriented and are intrinsically motivated and curious about the world when their basic psychological needs of autonomy, competence, and relatedness are met (Deci & Ryan, 1985). Skinner and Pitzer (2012) further stated motivation is intrinsic, and it is not acquired or lost. The psychological needs are universal across gender, age, race, and culture, and need to be facilitated in order for students to be motivated and, consequently, engaged in the classroom (Reeve, 2012). The primary tenet of SDT is that a student's level of intrinsic motivation is predicated on how well a student’s psychological needs of autonomy, competence, and relatedness are met by social contexts (Reeve, 2012).

Connell and Wellborn (1991) developed the Self-Systems Process Model (SSPM) based on SDT and focused on engagement rather than motivation (see Figure 2). In an SDT framework, satisfying the needs of autonomy, competence, and relatedness are a prerequisite for an individual to develop intrinsic motivation, whereas in the SSPM, satisfying needs of autonomy, competence, and relatedness leads to engagement.

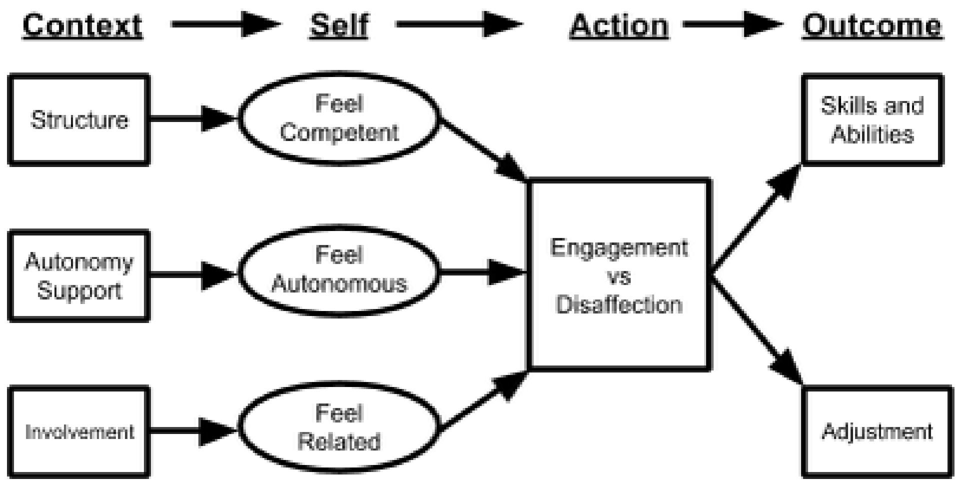


Figure 2. **Self-systems Process Model. Reprinted from "Competence, autonomy, and relatedness: A motivational analysis of self-system processes," J.P. Connell and J.G. Wellborn, 1991, *Self-processes in development*, p. 51. Reprinted with permission**

Motivation and engagement are closely related and are sometimes used interchangeably by researchers; however, they are distinctly different constructs as motivation is unobservable and private, and it is the drive or intent behind engagement, which is not private and is observable (Reeve, 2012). The linear SSPM, grounded in SDT (Skinner & Pitzer, 2012), identified that social context and environment (context) affect basic psychological needs (self) which in turn influence a student's level of engagement (action) and consequently achievement (outcome) (Reschly & Christenson, 2012; Skinner et al., 2008; Skinner & Pitzer, 2012). Students lacking motivation and engagement are not having their psychological needs met in the classroom (Reeve, 2012; Skinner, Furrer, Marchland, & Kindermann, 2008; Wonglorsaichon, Wongwanich, & Wiratchai, 2014), which has been shown to adversely impact student achievement (Roorda, Koomen, Split, & Oort, 2011). Students are more likely to be motivated/engaged and succeed if their needs for relatedness, competence, and autonomy

are met (Reyes, Brackett, Rivers, White, & Salovey, 2012) by involvement in activities that are hands-on, heads-on, project-based, relevant to student lives, progressive, and interdisciplinary (Skinner & Pitzer, 2012) and by providing classrooms high in emotional climate (Reyes et al., 2012). Missing in prior research based on SDT and the SSPM is the use of student growth as an outcome.

Reeve (2012) has provided evidence that the flow of influence within the self-systems process model is more bidirectional with feedback loops. It is not just social context that influences motivation and engagement, but also the result of motivation and engagement influencing social context. Reschly and Christenson (2012) also found support for the model as they found engagement mediated the effect of psychological needs on student achievement. Motivation and terminology used to describe motivation will not be addressed in this research; however, it is assumed to go hand in hand with engagement.

Teacher-student Relationship - The Context

According to the SSPM proposed by Connell and Wellborn (1991), context is a distal process that influences student outcomes and is part of the context of the classroom. Various factors influence outcomes or student achievement, in schools, which was the focus of Hattie's (2009) synthesis of meta-analyses. Hattie identified 138 variables that impact student achievement in schools, with effect sizes ranging from 1.44 to -.34. Hattie categorized the 138 variables into six groups that influence achievement and included the student (d = .40), the home (d = .31), the school (d = .23), the curriculum (d = .45), the teacher (d = .49), and approach to teaching (d = .42) with Cohen's effect sizes listed. Hattie, using a bar of an effect size of d = .4, concluded

teachers had the greatest contribution to student learning within the classroom. The results of Hattie's work indicated that what teachers do in their classrooms matters.

The *American Heritage College Dictionary* defines relationship as "a particular type of connection between people related to or having dealings with each other" (p.1152). Although no clear definition of the teacher-student relationship (TSR) has been established, many researchers have demonstrated there are many characteristics of TSRs, and that TSRs impact student outcomes.

Characteristics found to describe TSRs include, but were not limited to, teacher involvement (Fredricks et al., 2004), trustworthiness, accepting and respectful (Hughes, Wu, Kwok, Villarreal, & Johnson, 2012), warmth and empathy towards student needs (Wentzel, 2002), friendliness (Rickards & Fisher, 1997), and availability (Smart, 2014). Pianta's (2001) Student Teacher Relationship Scale utilizes the factors of closeness, dependency, and conflict as measures of the TSR. TSRs were influenced by student, teacher, and environmental characteristics and were the result of interplay between student, teacher, and environment, each, influencing each other (Rudasill & Rimm-Kaufman, 2009). In their review, Rudasill and Rimm-Kaufman, found that positive TSRs allowed students to use social skills to work through challenges, provided safety nets for students at academic risk, and promoted positive feelings towards school.

Cornelius-White (2007) found TSRs had a significant positive impact on all student outcomes with large effect sizes, with student outcomes consisting of measures such as grade point average, perceived achievement, IQ, attendance, behavior, measures of creativity, self-esteem, and social adjustment. Positive TSRs have been associated with increased motivation and academic achievement (Klem & Connell, 2004; Wilkins,

2014). Wubbels and Levy (1993) found that 70% of variability in student achievement was due to student perception of interpersonal teacher behavior, which influences the TSR. Students, parents, and principals identified teachers and their relationships with students as the main influence of student achievement (Hattie, 2009). From a Self-Systems Process Model perspective, a strong student perception of a positive TSR fulfills the underlying basic psychological needs of autonomy, competence, and relatedness of a student (Hughes et al., 2012).

Basic Psychological Needs - The Self

The perception of contexts in which an individual is situated influences one's sense of self and satisfaction of the psychological needs of autonomy, competence, and relatedness (Connell & Wellborn, 1991). Fulfilling needs of autonomy, competence, and relatedness can be fostered by teachers and social contexts in the classroom (Fried & Konza, 2013).

Autonomy, an individual's desire to act in accordance with one's self (Stroet et al., 2013) based on values and needs (Opdenakker & Minnaert, 2014) is promoted in classrooms by providing for student choice, implementing lessons based on student interests, having respect for student ideas and opinions, and providing constructive feedback (Fried & Konza; Stroet et al., 2013).

Competence refers to an individual's beliefs about one's capabilities and sense of effectiveness in dealing with the social context (Opdenakker & Minnaert, 2014). Individuals need to feel they are capable and can become more capable (Stroet et al., 2013) and can be successful in challenging activities (Deci & Ryan, 2009).

Relatedness is defined as the need to establish and maintain lasting relationships with others and to be cared for by others while also caring for others (Opdenakker & Minnaert, 2014). Based on the literature review, belonging (Deci & Ryan, 2000a; Fredricks et al., 2004), connectedness (Furrer & Skinner, 2003), and involvement (Stroet et al., 2013) were all synonymous with relatedness. The need for relatedness is satisfied by providing warmth, support, and nurturance (Deci & Ryan) in the classroom, along with building personal conflict free relationships (Fried & Konza, 2013).

In the SSPM, satisfaction of autonomy, competence, and relatedness needs influences an individual's action in the form of engagement (Connell & Wellborn, 1991). Not meeting these psychological needs of students, according to the model, will lower students' engagement levels and, consequently, have a negative affect on outcomes, such as achievement.

Engagement - The Action

Engagement has been shown by a number of researchers to impact student achievement, with the consensus being that the more engaged students are, the higher they will achieve (Duffield, Wageman, & Hodge, 2013; Fredricks et al., 2004; Klem & Connell, 2004; Sever et al., 2014; Wonglorsaichon et al., 2014). Engagement is a proximal process and a direct pathway to learning and achievement (Lawson & Lawson, 2013; Skinner & Pitzer, 2012). At the high school level, it has been estimated that 40% to 60% of students are not fully engaged, but are bored (Conner & Pope, 2013), with 66% of high school students reporting being bored in class every day by Yazzie-Mintz (2010) using the High School Survey of Student Engagement at 103 schools in 27 U.S. states.

Engagement has been defined in various ways by researchers with differing numbers of dimensions. For the purpose of this research, engagement consisted of three

dimensions which included behavioral, cognitive, and emotional/affective engagement. Behavioral engagement is defined as observable actions of students (Fredricks et al., 2004; Mahatmya, Lohman, Matjasko, & Farb, 2012) and consists of behaviors such as conduct, classroom participation, paying attention, working on tasks (Wang, Bergin, & Bergin, 2014). Cognitive engagement is identified by a student's psychological investment and willingness to put in effort, and it consists of behaviors such as spending time thinking about and reflecting on ideas and how to solve problems (Skinner & Pitzer, 2012; Wang et al.). Emotional engagement is identified by a student's enjoyment of the atmosphere around him or her, interest in school, optimism and enthusiasm for school (Klem & Connell, 2004; Skinner & Pitzer, 2012; Wang et al.). Engagement in this research, overall, is summarized by Skinner and Pitzer (2012) through their definition, "constructive, enthusiastic, willing, emotionally positive, and cognitively focused participation with learning activities in school" (p. 22).

Some researchers include a fourth dimension of engagement, disaffection, which is withdrawal from learning tasks, lack of effort and concentration, boredom, anxiety, frustration, and going through the motions (Skinner et al., 2008). Disaffection was not included in this study.

Student Growth - The Outcome

All of the research cited thus far utilized various measures of student achievement such as GPAs, class averages, teacher test scores, and standardized status scores as measures of student outcomes. Student growth as determined by student growth percentiles is a new measure of student outcomes, and has not been included in prior research. The primary purpose of the growth model is to provide insight into student learning as a result of a specific school or teacher. Haertel (2013) stated, 60% of the

variance in achievement was accounted for by factors outside of a schools' and teachers' control; therefore, measures of student growth are set up to strip away those factors (Haertel, 2013), and attribute student learning to a teacher/principal/school (Betebenner, 2008; Doran, 2003). With this purpose in mind, use of SGPs in Georgia as a measure of teacher effectiveness eliminates many of Hattie's (2009) variables that focus on the home, the curriculum, the school, and the student because a teacher has no control over these, and they are not accounted for by SGPs. According to Huitt et al. (2009), what happens in the classroom between teacher and student was the most direct influence on student achievement in the classroom. Similarly, based on Hattie's synthesis of meta-analyses, the teacher and his or her approach to teaching had a significant influence on student achievement. Included in the review by Hattie was the work of Cornelius-White (2007), which focused on person-centered teacher variables and student outcomes, which had a Cohen's effect size of d = .72. The six variables of non-directive, empathy, warmth, encouragement of higher order thinking and learning, and adapting to differences, all pertaining to teacher-student relationships, had individual effect sizes of greater than d = .4 (Cornelius-White). Cornelius-White noted that in classrooms with positive teacher-student relationship, there was more engagement, more student initiated action, and greater student achievement.

Gaps in the Research

While there is copious research pertaining to how to improve student achievement, there is little research on classroom variables and how they impact student

growth as it pertains to student growth percentiles. A majority of the literature deals with which growth model, value added or student growth percentiles, is a better indicator of who or what influenced student growth and the validity of that measurement. Research on the impact of factors affecting student growth as determined by value added or SGP is rare, with the researcher finding only four dissertations on the subject (Cervoni, 2014; Craig, 2011; LeGeros, 2013; Simmons, 2006) and no published peer-reviewed research literature. Of the four dissertations, LeGeros was the only one to find a variable significantly correlated with student growth percentiles. LeGeros found that conditionally passing or fully passing the Massachusetts Teacher Education Licensing exam was significantly correlated with student growth percentiles at the elementary level. Growth measures, both value added and SGP, have been or will be implemented throughout the U.S. to make high stakes decisions about teachers; therefore, research needs to be undertaken to determine what factors improve student growth as measured by these tools. Logically, one would assume that by increasing achievement, student growth would be increased; however, based on the student growth percentile model used in Georgia and other states in the U.S., this may not be the case.

Summary

This research was driven by multiple gaps in the literature, the most significant being the lack of connection between classroom variables with student growth percentiles. While the literature indicated positive TSRs and higher levels of engagement were associated with higher levels of achievement, measures of achievement included self-reported GPA’s, teacher assessments and assigned grades, instruments created for research, and standardized assessments (Duffield et al., 2013; Fredricks et al., 2004;

Klem & Connell, 2004; Sever et al., 2014; Wonglorsaichon et al., 2014), all of which were not high stakes or were not used to make high stakes decisions on teacher effectiveness. These measures of achievement are status scores and are straightforward to interpret and have been around since education's inception, whereas growth scores are not straightforward, are relatively new to the educational landscape, and require a little more inspection to understand.

Using Connell and Wellborn's (1991) Self-Systems Process Model as a framework, no study was identified that evaluated the full model from context to self to action to outcome flow of the model except for the work completed by Connell and Wellborn, let alone the influence of the included factors on student growth percentiles. Archambault et al., (2009) examined the relationship between engagement (action) and student dropout (outcome), while Tian et al., (2015) examined the impact of teacher support (context) on competence, autonomy, and relatedness (self) with the outcome being student subjective well-being. Roorda et al., (2011) determined effect sizes of TSR (context) on engagement (action) and TSR (context) on achievement (outcome) separately, but did not include basic psychological needs (self) or evaluate the full model. Cornelius-White (2007) had similar limitations in that basic psychological needs were missing from the research.

Few studies exist that include the newly defined three dimensions of behavioral, cognitive, and emotional engagement at the classroom level. Prior studies typically included a differing number of dimensions such as Marks (2000), who used two dimensions that included behavioral and emotional engagement, and Fredricks et al., (2004), who used three dimensions that included behavioral, cognitive, and affective

engagement, and Reschly and Christenson (2006), who used four dimensions that included academic, behavioral, cognitive, and psychological engagement. Various instruments utilized in studies included items that measured both classroom and school level engagement, leading to false measures of classroom or school level engagement (Fredricks et al., 2004; Wang et al., 2014).

Therefore, the researcher proposes to add to the literature by addressing the deficiencies previously mentioned which include lack of full self-systems process model support, the multidimensionality of engagement, and factors influencing student growth. The intent of this research is to determine how TSRs (context), influence basic psychological needs (self), which influence engagement (action), and ultimately impact student growth scores/achievement status scores (outcome) using the full self-systems model from context to outcome by including the multidimensionality of engagement to include behavioral, cognitive, and affective engagement similar to Fredricks et al., Fredricks and McColskey (2012), and Wang et al., (2014) measured at the classroom level using the newly created Classroom Engagement Instrument developed by Wang et al., (2014). This research will build on prior findings in the research utilizing standardized assessment status scores as the dependent variable and then comparing the results with an identical methodological setup with student growth percentiles as the dependent variable.

Statement of the Problem

Prior to the implementation of the Teacher Keys Effectiveness System in the state of Georgia, student achievement was not considered in the evaluation of teacher

effectiveness. Student growth now plays a significant role in teacher evaluations in the state of Georgia, and identifying strategies teachers can implement in their classrooms to better support student growth is becoming increasingly important. Positive teacher-student relationships and high levels of engagement have been shown to improve student achievement as measured by student self-reported GPA's, teacher created assessments, assessments created for research, and standardized assessments; however, there has been no research on the influence of student growth as measured by student growth percentiles. Structural equation modeling will be used to understand how teacher-student relationships working through self-determined needs of autonomy, competence, and relatedness influence engagement and, consequently, influence student growth as measured using student growth percentiles with seventh and eighth-grade students.

Research Questions

The over-arching research question is as follows:

How does the teacher-student relationship influence student engagement as measured by Classroom Engagement Inventory (CEI) and student achievement as measured by student growth percentiles using a self-systems process model perspective?

Subquestions are as follows:

1. To what extent does the teacher-student relationship influence satisfaction of basic psychological needs which influence engagement and, consequently, influence student growth percentiles as compared to student status scores using an identical methodological setup (Context → Self → Action → Outcome)?

2. To what extent is the effect of teacher-student relationships on student growth percentiles invariant across population subgroups? (i.e. Low socioeconomic status

students versus high socioeconomic status students and White students versus non-white students)

3. To what extent does the teacher-student relationship influence level of student engagement (Context → Self → Action)?

CHAPTER II
REVIEW OF LITERATURE

Educator accountability is at the forefront of educational reform, with student growth counting as a significant portion of a teacher's overall evaluation. Peer reviewed research on factors that affect student growth, as determined by student growth percentiles, has not been identified, and only three dissertations on the subject with findings that are cause for concern. While there is copious research on how and what factors improve student achievement, it is unknown if there is a direct relationship with improving student growth, which prompted this research.

The hypothesized Self-System Process Model (SSPM), developed by Connell and Wellborn (1991), was used as the framework for this study. The review focused on the basis of the SSPM, self-determination theory, the components of the SSPM which include context, self, action, and outcome, status and growth models in general, and the specific student growth percentile model used in the state of Georgia. The purpose of the research was to investigate how student perceived teacher-student relationships influence basic psychological needs, engagement, and ultimately student growth using the SSPM proposed by Connell and Wellborn.

While there are other needs-based theories of motivation, such as that proposed by Maslow in his hierarchy of needs, Alderfer’s ERG theory, and McClelland’s acquired needs theory, the focus of this research was based on Deci and Ryan’s self-determination theory and basic psychological needs of autonomy, competence, and relatedness as the review shows is well supported empirically.

Self-Determination Theory

All individuals innately strive towards vitality, integration with others, and good health, and have instinctual needs that must be present to support their endeavors (Deci & Ryan, 2000b). Self-determination Theory (SDT) is an empirically-based hypothesized model of social motivation borne out of the need to understand and explain an individual's motivation (Deci & Ryan, 2009; Reeve, 2012). SDT posits that individuals are curious to the world around them and are intrinsically motivated to explore when underlying basic psychological needs (BPN) of autonomy, competence, and relatedness are satisfied (Reeve, 2012; Skinner & Pitzer, 2012), similar to the requirement of food and water for an individual to have proper physiological health (Deci & Ryan, 2000b). Deci & Ryan (2000a) identified intrinsic motivation, motivation from within oneself and endorsed by oneself, as a self-determined type of motivation in that it is autonomous. They further stated that intrinsically motivated students engage without feeling coerced or controlled by an outside entity for their own sake.

Psychological needs include autonomy, which is perceived choice, and the ability for students to make important decisions regarding their learning (Klem & Connell, 2004), competence, which is being effective at some task or skill, and relatedness, which is establishing bonds with others such as peers, teachers, and school that are caring and nurturing (Skinner & Pitzer, 2012). All three nutriments, autonomy, competence, and relatedness, are equally important because a deficit in one can cause lower levels of psychological functioning and experience (Connell & Wellborn, 1991; Ryan & Deci, 2001; Deci & Ryan, 2002) as Tian et al., (2014) found the BPNs to be highly related.

Deci and Ryan (2002) stated that BPNs “specify innate psychological nutriments that are essential for ongoing psychological growth, integrity, and well-being” (p.229),

and that BPN satisfaction is the underlying motivational mechanism that energizes and drives people's behavior. The way in which need satisfaction promotes individual development is theorized to be invariant across age, gender, and culture (Reeve, 2012; Ryan & Deci, 2001). While individuals of different culture, age, and gender may satisfy BPNs in different ways, the individuals will benefit from having BPNs fulfilled (Deci & Ryan).

While motivation and engagement are distinctly different due to the fact that motivation reflects underlying energy and intention, while engagement reflects action and doing (Lawson & Lawson, 2013; Skinner & Pitzer, 2012; Reschly & Christenson, 2012), engagement is a manifestation of intrinsic motivation (Skinner, Kindermann, & Furrer, 2009; Wonglorsaichon et al., 2014), and motivation research typically includes an action component that shares characteristics with engagement. Deci and Ryan (2009), the fathers of SDT, supported this claim when stating, “intrinsic motivation concerns active engagement with tasks that people find interesting and that, in turn promote growth” (p. 233). They further showed the relationship between the two in this statement:

> This active engagement, this involvement and commitment with interesting activities, requires the nutriments of need fulfillment, and, indeed, people will become more or less interested in activities as a function of the degree to which they experience need satisfaction while engaging in those activities (p. 233).

Intrinsic motivation is not observable because it is an internal private process that is an antecedent to engagement, which is observable (Reeve, 2012). Skinner and Pitzer, (2012) alluded to this in stating, “Engagement refers to energized, directed, and sustained action, or the observable qualities of students’ actual interactions with academic tasks” (p. 24). Reschly and Christenson (2012) clarified that it is generally accepted that

motivation and engagement are linked and influenced by context and are unique to individuals.

Self-Systems Process Model

Connell & Wellborn (1991) developed the Self-Systems Process Model (SSPM) based on SDT (Figure 2) and included engagement rather than intrinsic motivation as the result of nourishment of psychological needs. According to this model, "the objective self is the individual's appraisal of how competent, autonomous, and related he or she feels within and across particular contexts. These appraisal processes are referred to as self-system processes" (Connell & Wellborn, p. 52) which arise out of interaction between social contexts. Similar to SDT, the SSPM requires satisfaction of the basic psychological needs of autonomy, competence, and relatedness, but does so through social context, with the classroom, teacher-student interactions, and the teacher-student relationship representing the context in this research. Self-system processes develop out of the interaction between psychological needs and social context; aspects of social context that influence basic psychological needs are of greatest importance as they drive action and outcome according to the model. Social interactions with peers and/or teachers within the classroom (context) either support or hinder psychological needs (self) which influence an engagement (action), which in turn influences skills, abilities, and adjustment (outcomes). It is an individual's experience of social context that contributes to the development of the self-system. Connell and Wellborn noted that a poor person-environment fit will inhibit psychological well-being and the self-system. They also identified that an individual's perception of social context is paramount as individual perception drives the self-system processes.

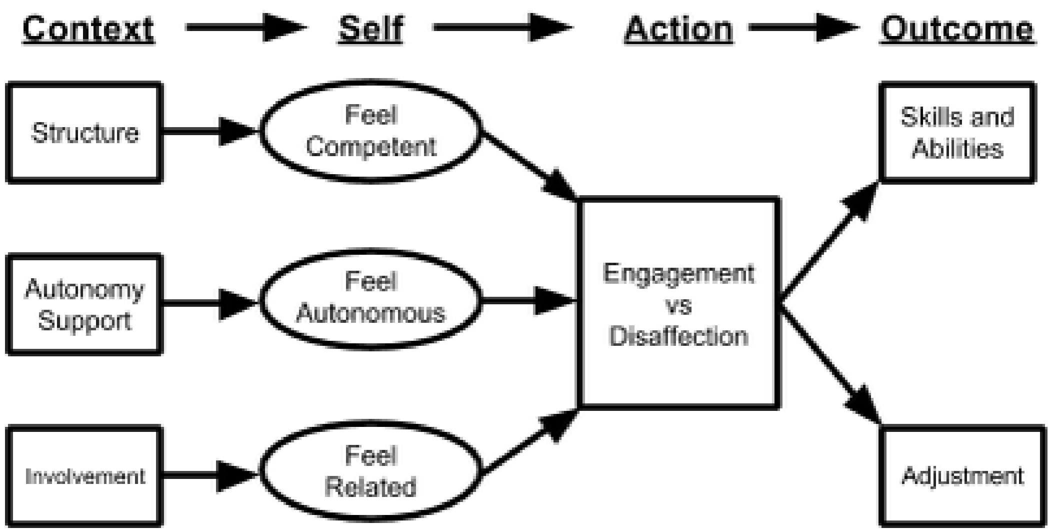


Figure 2. Self-systems Process Model. Reprinted from "Competence, autonomy, and relatedness: A motivational analysis of self-system processes," J.P. Connell and J.G. Wellborn, 1991, *Self-processes in development*, p. 51. Reprinted with permission

In their research, Connell and Wellborn (1991) identified that engagement mediated the effect of BPNs on outcomes. In multiple studies of third to sixth graders, Connell and Wellborn provided support for the SSPM in Figure 2. Utilizing path analysis, they found a direct relationship between competence, autonomy, and relatedness and teacher rated engagement along with a direct relationship between teacher rated engagement and student achievement test scores. In the same sample, Connell and Wellborn found significant correlations between relatedness and teacher ratings of engagement, yet there were no correlations to achievement, highlighting the mediational effect of engagement in the model. Reschly and Christenson (2012) found support in their research for the mediational effect of engagement on psychological needs on student achievement.

The original SSPM was linear in nature, identifying that social context affected psychological needs and motivation, which then influenced engagement and,

consequently, achievement (Connell & Wellborn, 1991; Reschly & Christenson, 2012; Skinner et al., 2008; Skinner & Pitzer, 2012). However, Reeve (2012) provided evidence that the process is more bidirectional with feedback loops. It is not just social context that influences motivation and engagement, but also the result of motivation and engagement influencing social context.

As motivation and engagement are linked, a student's level of engagement wihtin the classroom is a reflection of how well basic psychological needs of autonomy, competence, and relatedness are met within social context (Hughes et al., 2012; Stroet, Opdenakker, & Minnaert, 2013), with motivation and engagement operating optimally when psychological nutriments are present (Deci & Ryan, 2000b). Intrinsic motivation is not acquired or lost, but can decline when students' psychological needs are not being met by schools and teachers (Skinner & Pitzer, 2012), because intrinsic motivation is a reflection of satisfaction of psychological needs (Deci & Ryan, 2000b). Students lacking motivation and engagement have likely not had their psychological needs met in the classroom, and therefore, have had lower levels of engagement (Reeve, 2012; Skinner et al., 2008; Wonglorsaichon et al., 2014) and, consequently, lower levels of achievement (Roorda et al., 2011). The greater extent psychological needs are met, the greater driving force students will have; the less the needs are met, the less motivation they will have (Reeve, 2012).

Students are more likely to be motivated/engaged and successful if their needs for relatedness, competence, and autonomy are met in the classroom (Reyes et al., 2012). The needs of students can be satisfied through social contexts; however, students interpret and react differently to social contexts due to their unique identities and

experiences (Skinner & Pitzer, 2012). Psychological needs can be supported by providing activities that are hands-on, heads-on, project-based, relevant to student lives, progressive, and interdisciplinary (Skinner & Pitzer, 2012), and by providing classrooms high in emotional climate (Reyes et al., 2012).

Model support**.** Roorda et al. (2011) completed a meta-analytic review of the literature on affective TSR, engagement, and achievement which included 99 studies, 129,423 K-12 students, and 2,825 teachers from five continents from 1990 to 2011. Only studies with engagement and achievement as dependent variables and TSR as independent variables were included in the analysis. While Roorda et al., did not specifically address the full SSPM, context, action, and outcomes were studied. Prior research indicated the quality of teacher-student relationships had an impact on engagement and academic achievement with poor relationships having more of a negative impact than good relationships having a positive impact (Roorda et al.). Relationships were not straightforward and may have been affected by student characteristics such as age, gender, ethnicity, socioeconomic status (SES) and teacher characteristics such as gender, ethnicity, and experience (Roorda et al.). The analysis focused on positive and negative affective dimensions of person-centered teacher behaviors. Similar to Hattie, (2009), effect sizes greater than or equal to .4 were considered large and of great importance. Effect sizes of .25 to .4 were considered medium to large, and effect sizes of .10 to .25 as small.

Due to findings in other research that engagement has been found to act as a mediator between TSR and achievement, Roorda et al., (2011) hypothesized TSR would have a stronger association with engagement than achievement. Including all studies

with both similar and different informants, for random and fixed effects studies, effect sizes indicated that positive affective TSRs had positive associations with engagement and achievement with the latter smaller, and negative affective TSRs had negative associations with engagement and achievement with the latter smaller (see Table 1, Roorda et al.).

Table 1

Effect sizes of TSR on engagement and achievement for fixed and random effects studies

	Positive TSRs and engagement	Negative TSRs and engagement	Positive TSRs and achievement	Negative TSRs and achievement
Fixed	.39	-.32	.16	-.15
Random	.34	-.31	.16	-.18

The level of association between TSR and engagement varied depending on informants. Using the same informant yielded larger effect sizes for the effect of both positive and negative TSR on engagement but yielded smaller effect sizes for the effect of both positive and negative TSR on achievement (see Table 2). The influence of TSR on engagement was larger when the same informants were used, possibly attributed to shared variance of using same informant (Roorda et al., 2011) similar to the findings of Reyes et al. (2012).

Table 2

Effect sizes of TSR on engagement and achievement for informant type

	Positive TSRs and engagement	Negative TSRs and engagement	Positive TSRs and achievement	Negative TSRs and achievement

Table 2

Effect sizes of TSR on engagement and achievement for informant type

Same Informant	.41	-.42	.14	-.13
Different Informant	.23	-.30	.17	-.19

Roorda et al., (2011) included a model similar to the SSPM, but did not investigate the full path from context to self to action. Instead, the researchers investigated the direct relationship of context to action and context to outcome. Overall, TSR had greater ties to engagement than to achievement, which is supported by SDT in that TSR is more proximal to autonomy, competence, and relatedness than academic achievement. “In line with the self-determination theory, the smaller associations with achievement seem to suggest that the effect of TSRs on achievement runs partly via engagement” (Roorda et al., 2011, p.516).

Furrer and Skinner (2003) completed research on relatedness and resulting student engagement and performance in a sample of 641 third to sixth grade students from one school district. Behavioral and emotional engagement was measured from both the teacher and student perspectives along with student reported levels of relatedness. Student reports of engagement and relatedness were highly correlated, while teacher rating of student engagement and academic performance were related. The latter finding was a possible indication that students’ engagement levels played a role in assignment of student grades by the teacher (Furrer & Skinner). Correlations between teacher and student reports of emotional engagement were lower than teacher and student reports of behavioral engagement. Skinner and Furrer attributed this to behavioral engagement

having been more observable outward actions that were easily identifiable by teachers, as compared to emotional engagement, which is an internal process and difficult for teachers to perceive.

The work of Furrer and Skinner (2003) supported the SSPM in that relatedness (self) influenced engagement (action), which influenced achievement (outcome). Students who felt connected had more positive emotions, energy, interest, and willingness, providing the energy inputs (engagement) in the schooling process. While Furrer and Skinner concluded that students relatedness is a key component needed in the classroom for student success, generalizations may be limited because the sample population was 95% Caucasian, and no high school students were included.

Hughes, Luo, Kwok, & Loyd, (2008) hypothesized in a three-year longitudinal study, TSR in year one would affect achievement in year three and be mediated by engagement in year two. Participants consisted of 671 ethnically diverse students from three Texas schools. Measurements included teacher perception of TSR, Woodcock-Johnson III test of achievement, teacher report of effortful engagement, and teacher rated conduct engagement. Hughes et al., indicated teachers' rated levels of engagement in year two, mediated the effect of TSR quality in year-one on achievement in year-three, with math results being more robust than reading across all students.

Similar in structure to prior research by Hughes et al., (2008), Hughes et al., (2012) collected student reports on warmth and conflict, teacher-rated engagement, student-rated perceived academic competence, and academic achievement as measured by the Woodcock-Johnson III test of achievement in a sample of 690 ethnically diverse students with the focus more on student perception. The purpose of their longitudinal

research was to determine if their hypothesized path model of TSR influencing teacher rated engagement, which then influenced math and reading achievement across three years, was supported.

Student perceived levels of warmth decreased from third to fifth grade for both males and females and different racial groups prior to the transition of the middle school. Levels of warmth influenced achievement, even though warmth levels as perceived by students dropped. Hughes et al., (2012) hypothesized that larger class sizes, focus on instruction instead of relationships, and less time in small groups all influenced students' perception of teacher warmth. Student rated conflict lowered teacher rated engagement levels in years two and three which, consequently, lowered achievement levels in years two and three. Student-rated warmth did not affect engagement, but did affect student-rated competence.

While the samples in both Hughes et al., (2008) and Hughes et al, (2012) were ethnically diverse, the samples were only of students who scored at or below the median level on standardized assessments of literacy, limiting generalizability to the general population. While other research indicated that academically at risk students benefit from positive TSR with higher levels of engagement and achievement, Hughes et al, (2012) found adversity indirectly influenced achievement more than support, which was likely due to the selected sample of participants being the lower achieving students. Another issue with the research was use of Woodstock-Johnson Tests of Achievement as a measure of achievement, as it is a nationally normed test that did not align with classroom standards. While it did provide a measure of achievement, it may have provided distorted picture of student achievement as it was not a measure of achievement

in participants' classrooms. Both Hughes et al., (2008) and Hughes et al., (2012) utilized a simplified SSPM in which psychological needs (self) was removed.

Stroet et al., (2013) conducted a meta-analytic review of 71 papers to determine if needs-supportive behaviors impacted motivation and engagement and supported SDT. Research in the review was included if the topic dealt with motivation, engagement, and early adolescence. Stroet et al., concluded that student perceived autonomy and structure supported student engagement. The flow of influence appeared to go through the psychological needs of competence and relatedness to engagement. Stroet et al., also concluded that students who perceived their teachers to be more involved, were more involved with students had students that were more engaged. These studies indicated that student-perceived teacher support was positively associated with motivation and achievement.

Contrary to the findings based on student perceptions, most studies of teacher-rated needs supportive environment found little to no association with motivation and engagement, representing a problem, as there is no continuity in teacher thought and student perception. Stroet et al., (2013) theorized student perceived needs support is individualized and a more accurate reflection of a student's psychological needs. Findings may be misaligned due to the possibility that student perception is easier to measure than is concrete, observable behaviors that represent those perceptions. In SDT, student perceptions of psychological needs drive motivation and engagement, not what teachers and administrators believe students psychological needs to be (Deci & Ryan, 2009).

Sakiz, Pape, and Hoy (2012) conducted research at four Midwest middle schools with 317 seventh and eighth-grade students in 39 math classrooms using structural equation modeling. Measurements included student perception of teachers' affective behaviors, their own sense of belonging, academic enjoyment, self-efficacy, and effort. While the hypothesized structural model was complex, had only reasonable fit, and contained one academic area with a mainly white population, Sakiz et al., found higher teacher support led to higher student self-report of belonging, enjoyment with material, and greater effort. Similar to Hughes et al., (2008, 2012), basic psychological needs were not included. There is general consensus that classroom structure, autonomy, and caring supportive relationships support student engagement (Conner & Pope, 2013; Fredricks et al., 2004) along with social context (Skinner et al., 2008) and teacher student relationships (Roorda et al., 2011), all of which are facilitators of engagement as theorized by self-systems process.

A consistent theme identified thus far is that student perception of the TSR, basic psychological needs, and engagement, from a SSPM perspective, is a better assessment of reality than teacher perception. According to Stroet et al. (2013), student needs are highly individualized and a more accurate reflection than teacher perception of student needs. According to Fisher and Rickards (1998), teachers favor their own behaviors and supportive actions more favorably than do students. Burniske and Melbaum (2012) stated that student ratings of teachers have been found to be consistent from year to year, students' ratings were as valid and reliable as adults', and students were able to discriminate differences in characteristics of teachers, especially when dealing with warm, caring interpersonal relationships.

Context (TSR) and self (BPNS) as perceived by individuals within the SSPM, drive the self-systems process (Connell & Wellborn, 1991). Skinner and Pitzer (2012) supported student perception as driving the model, and indicated that it was the result of students having different perceptions of reality based on their life's experiences (Skinner & Pitzer, 2012).

Teacher-student relationship**.** Three major sources of influence on students were peers, teachers, and parents (Martin, 2014). Teachers influenced students' emotional, social, and academic experiences at school due to high amount of interaction with students (Wilkins, 2014). Wubbels and Levy (1993) stated that 70% of variability in student achievement, and 55% of variability in student attitudes, was due to student perception of interpersonal teacher behavior. Positive TSRs have been associated with increased motivation and academic achievement (Birch & Ladd, 1997; Klem & Connell, 2004; Wilkins, 2014). SDT posits that in order for students to be intrinsically motivated and engaged in school activities, three psychological needs of relatedness, competence, and autonomy must be met (Deci & Ryan, 2000a). Autonomy, perceived competence, and belonging can be fostered by teachers and social contexts in the classroom (Fried & Konza, 2013), with the result of the interaction having an impact on intrinsic motivation, cognition, and well-being (Deci & Ryan, 2009). Specifically, teachers must care for and be genuinely interested in the student, set clear rules and expectations, and have consequences that are applied uniformly, provide choice in all aspects of education, and relate the work to student interests (Wilkins, 2014).

Teacher-student relationships (TSR) were influenced by student characteristics, teacher characteristics, and characteristics of the environment that both the student and

teacher were a part of through bidirectional interactions (Rudasill & Rimm-Kaufman, 2009). Rudasill and Rimm-Kaufman went on to state that TSRs are complex and are a result of interplay between student and teacher characteristics and daily interactions, which were constantly influencing each other and the perception of each other.

Rudasill and Rimm-Kaufman (2009) surmised that teacher perception of student characteristics could influence the quality of teacher-student relationship in that students who were attentive, sat still, and contributed to the class contributed to positive teacher relationships. Teachers who were trustworthy, accepting and respecting of all students, and available promoted autonomy, competence, and relatedness (Hughes et al., 2012). Teacher classroom control, propensity for angering quickly, and unwillingness to listen were negative teacher behaviors, whereas availability, approachability, and individualized attention were positive teacher behaviors according to students (Smart, 2014). Rickards and Fisher (1997) found the best teachers, as identified by students, possessed leadership, friendliness, and understanding according to questionnaire of teacher interaction results. Rickards and Fisher further stated that when students perceived greater leadership and helpful/friendly behaviors from their teachers, students had more favorable attitudes toward the class. Students identified caring and supportive teachers as those who promoted respectful and democratic interactions, had expectations contingent of individual abilities, are were warm and empathetic to student needs, and provided constructive feedback (Wentzel, 2002). High quality student-teacher interactions were categorized by Smart (2014) as consistent, stable, respectful, and fair and included rich dialogue and instructional exchanges between teacher and student and perceived emotional support.

Positive TSRs were marked as low in conflict and dependency by teachers, and high in closeness, respect, and caring from teachers seen as source of security by students (Rudasill & Rimm-Kaufman, 2009). Increased class sizes and change in teacher behaviors, which occurs as student progress in grade levels, were perceived by students as less support (Smart, 2014). Aggression, as marked by high levels of conflict and low levels of closeness and withdrawal, were negative predictors of TSR (Rudasill & Rimm-Kaufman, 2009). In their review, Rudasill and Rimm-Kaufman found that positive TSRs allowed students to use social skills to work through challenges, provided safety nets for students at academic risk, and promoted positive feelings towards school.

The classroom climate of middle schools was different from elementary climate as classes were larger, students had more teachers, and there was less parental involvement (Smart, 2014). Middle school structure was a poor person-environment fit that hindered relationship development due to students having multiple classrooms, larger class sizes, more standardized testing, and more curriculum to be covered, all of which provided fewer opportunities for students and teachers to connect with each other (Reddy, Rhodes, & Mulhall, 2003). Smart (2014), in her research of middle school science students, indicated teachers were more controlling, exhibited less nurturing behaviors, and provided fewer opportunities for student choice and decision making. According to Smart, students were keenly aware of their teachers' uncooperative behaviors, impatience, and frustration when they struggled with their work. Students were more motivated to learn middle school science when the teacher was willing to listen and be patient with them (Smart). Students who lacked positive relationships with

their teachers were more likely to avoid school, to feel lonely, and to display low levels of academic competence (Rudasill & Rimm-Kaufman, 2009).

Rudasill and Rimm-Kaufman, (2009) in their study of 819 children from first grade using the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development data set, parents' measure of child temperament, observational data of classroom interactions, teacher reported relationship quality, and structural equation modeling, found that shyness, effortful control, and gender predicted the quality of TSRs. Lower levels of shyness were associated with higher levels of conflict and closeness, and lower levels of effortful control were associated with higher levels of conflict. Males' relationships with teachers were marked by conflict, whereas female's relationships were marked by closeness (Rudasill & Rimm-Kaufman, 2009).

To understand how teachers defined a good TSR, Wilkins (2014) utilized a mixed methods study of a large urban high school with a large number of low SES students that included 103 teacher survey responses and six teacher interviews. From factor analysis of teacher survey items, three factors were identified that included students showing interest in school and school work, respect for teachers and the rules, and positive social behaviors such as having conversations with teachers outside of the classroom. Four themes that arose from teacher interviews, similar to the factors previously identified, included having respect not only for the teacher, but also the classroom and school, trying hard to do the work, talking with teachers about topics other than academic subject areas, and having a sense of humor, which was not an original factor.

While many prior studies point to lower quality relationships at the high school level (Skinner et al., 2008; Smart, 2014), teachers in Wilkins (2014) study identified relationships with their students as important in their positions because good relationships were a way to combat discipline problems and motivate students in the classroom, which helped with instruction. One teacher noted, “If they feel comfortable in a non-threatening environment, they will perform better--it’s true!” (Wilkins, 2014, p. 66). While Wilkins found that relationships were important, relationships by teachers were predicated on students’ effort. Multiple teachers stated that they would put forth less effort for students who made no effort, which is what Rudasill and Rimm-Kaufman (2009) alluded to in stating that daily interactions constantly influenced each other and the perception of each other. Generalization of Wilkin’s findings may be problematic, because teacher survey completion was voluntary with an 18% completion rate. It was highly possible that teachers with good TSRs responded at a higher rate than teachers who did not have good TSRs.

Hamre and Pianta (2001) identified three dimensions to the TSR, according to teacher reports, which included closeness, dependency, and conflict and were found to be invariant across age, ethnicity, and socioeconomic status. Birch and Ladd (1997) defined closeness as the degree of warmth and open dialogue between student and teacher, dependency as the amount of reliance on the teacher, and conflict as lack of rapport and as having friction between teacher and student. These indicators, that are teacher reported, are part of the Student Teacher Relationship Scale (STRS) developed by Pianta (2001) which has been utilized in many studies of TSRs (Birch & Ladd, 1997; Birch & Ladd, 1998; Rudasill & Rimm-Kaufman, 2009). Birch and Ladd (1997), in their research

of 206 kindergarten students and their teachers using the STRS, found that level of closeness was significantly correlated to student academic performance on the Metropolitan Readiness Test. The result, according to Birch and Ladd, may have been due to students being able to utilize their teacher as a source of support, which allowed them to benefit from classroom activities. Hamre and Pianta, in their research of 179 students in a small school district, found that negative TSRs were a significant predictor of academic outcomes. They surmised that students who experienced lower levels of closeness and higher levels of dependency and conflict were less motivated to succeed. Pianta, Hamre, and Allen (2012), in later work, suggested that closeness, dependency, and conflict were part of the classroom climate, which is measured along a continuum. A positive climate is marked by warmth and caring between teachers and peers, whereas a negative climate is marked by conflict in which there is humiliation, yelling, and rejection between teachers and peers.

The Network of Relationships Inventory (NRI) was developed to measure relationship characteristics across a range of relationships that included siblings, parents, other significant adults, and teachers and is self-reported by students rather than reported by an outside observer (Furman & Buhrmester, 1985). The Network of Relationships Inventory - Relationship Quality Version (NRI-RQV) assesses the two second order factors of closeness and discord through multiple first order factors, all of which are measures of relationship quality similar to Pianta's (2001) STRS. The NRI-RQV differs from the STRS in that it is from the student perspective and is appropriate for use in children eleven and up, whereas the STRS is from a teacher perspective.

Basic psychological needs. Stroet et al., (2013) defined autonomy as a student's desire to act in accordance with one's self, which aligns with Deci and Ryan (2009) defining autonomy as regulating one's own behaviors. A student internally wants to act on his or her own accord based on needs and values, and with no pressure from outside influences (Opdenakker & Minnaert, 2014). Classrooms exhibiting autonomy-supportive characteristics were student-centered, allowed a student to have a voice by providing student choices, fostered relevance to student interests, showed respect, provided constructive criticism, and utilized informational language (Fried & Konza, 2013; Stroet et al., 2013). Students having no choice, perceiving the curriculum to be irrelevant, lacking respect for the teacher, or the teacher sending signals of harsh criticism and controlling language were characteristic of classrooms that did not promote autonomy (Stroet et al.).

Competence was a sense of effectiveness in dealing with the social environment according to Opdenakker and Minnaert (2014). Individuals need to feel they are capable and can become more capable (Stroet et al., 2013) and engage in challenging activities with success (Deci & Ryan, 2009). A student's sense of competence prepared them for the challenges of school work and provided energy for learning (Opdenakker & Minnaert).

Words synonymous with relatedness included belonging (Deci & Ryan, 2000; Fredricks et al., 2004) connectedness (Furrer & Skinner, 2003), and involvement (Stroet et al., 2013). Relatedness was defined as the need to establish and maintain lasting relationships with others and to be cared for by others while also caring for others (Opdenakker and Minnaert, 2014). Similar to relatedness, Fredricks et al., (2004) defined

belonging as being accepted, valued, and included. Involvement was defined as the need to maintain stable interpersonal relationships that will last and be conflict free, be connected to others, and to belong (Stroet et al.). Similar to relatedness, classroom emotional climate (CEC) consisted of the quality of social and emotional interactions between teachers, students, and the classroom (Reyes et al., 2012). Reyes et al. noted high levels of CEC were marked by classrooms which were sensitive to student's needs, provided caring and nurturing relationships with little sarcasm and harsh disciplinary action, and were an open classroom with respectful interactions that focused on student interests.

The need for relatedness was satisfied by providing warmth, support, and nurturance (Deci & Ryan, 2000a). Belonging was fostered by building personal relationships within the classroom (Fried & Konza, 2013). Students that reported more connectedness to teachers reported more involvement with activities and had more positive emotions, whereas children low in connectedness did not feel emotionally attached to peers, teachers, and parents, and were more likely to become bored and alienated, further withdrawing from school activities (Furrer & Skinner, 2003). Students that reported not feeling important and/or being ignored by teachers were less happy and experienced more boredom while at school (Furrer & Skinner). When relatedness is provided for, students adopt and internalize external behaviors, values, and beliefs of those around them, which supports engagement in schools (Opdenakker & Minnaert, 2014). Consistent evidence showed the higher levels of relatedness, connectedness, and belonging to community were associated with higher behavioral and emotional engagement (Fredricks et al., 2004). Students that reported greater levels of relatedness

worked harder, had more positive affect, and had greater academic success (Furrer & Skinner, 2003).

Engagement**.** Engagement is a relatively new idea in education (Reschly & Christenson, 2012), and it has been estimated that 40% to 60% of high school students are not fully engaged in the classroom, do not complete their work, and report being bored (Conner & Pope, 2013). Yazzie-Mintz, (2010) reported that 66% of high school students reported being bored in class every day. Conner and Pope (2013) in their review, cited lack of challenge, uninteresting material and content, and lack of interaction with the teacher as factors found to cause lack of engagement and boredom. Engagement is a proximal process, direct pathway to learning and achievement (Lawson & Lawson, 2013; Skinner & Pitzer, 2012), and predicts achievement levels (Skinner et al., 2008). Since the inception of this metaconstruct, there has been little consensus on the number of dimensions and the definition of each in the literature except for a base of participatory behavior and some affective components (Reschly & Christenson, 2012) with recent research focusing on three dimensions. Moreover, student engagement is malleable and can be influenced by social contexts, culture of the classroom, tasks (Lawson & Lawson, 2013; Mahatmya et al., 2012; Skinner & Pitzer, 2012), parents, peers, and teachers (Wonglorsaichon et al., 2014), with many educational interventions attempting to change a student's level of engagement (Fredricks et al., 2004; Sever, Ulubey, Toraman, & Türe, 2014;). Indicators of engagement are actions students take and are aspects of engagement to be measured such as time on task or participation in discussions (Skinner & Pitzer, 2012). Many interventions located in the *What Works Clearinghouse* pertain to addressing engagement levels of students (Reschly & Christenson, 2012). The

complexity of the construct of engagement requires further elaboration in order to clearly understand its role in student achievement.

Definition. Wang et al, (2014) defined engagement as “a student’s active involvement in classroom learning activities” which includes “attention, interest, investment, and effort students expend in the work of learning” (p.517) while Wonglorsaichon et al., (2014) defined engagement as “students’ expression of opinions or attitudes and behaviors” (p. 1749). Characteristics exhibited by engaged students include participation in class activities, being attentive, showing interest in the class and learning, and being effortful (Reyes et al., 2012). The majority of recent literature on engagement identified engagement as multidimensional, with three distinct types which include affective or emotional, cognitive, and behavioral engagement (Mahatmya et al., 2012; Reschly & Christenson, 2012; Wang et al; Wonglorsaichon et al.) with varying levels of each type of engagement in different classrooms (Wang et al).

A relatively new dimension, disengagement, also known as disaffection, has been included by some researchers as a form of engagement (Wang et al., (2014). Skinner & Pitzer (2012) added that disaffection, was more than a low level or absence of engagement, and in fact, was a willful withdrawal from learning tasks, lack of effort and concentration, and boredom.

Affective engagement was characterized by positive and negative reactions to aspects of schooling (Mahatmya et al., 2012) such as positive emotions linked to peers, classrooms, teachers, and the school (Wonglorsaichon et al., 2014). Other synonymous terms used to identify emotional engagement included enjoyment of atmosphere, interest in school, optimism, and enthusiasm for school (Klem & Connell, 2004; Skinner &

Pitzer, 2012; Wang et al, 2014). It is the reaction and attitude towards school, teachers, students, and the environment that is tied to students' willingness to work in school. Identification with school "refers to students' affective reactions in the classroom, including interest, boredom, happiness, sadness, and anxiety" (Fredricks et al., 2004, p.63) and deals with students' social and emotional attachments to school (Lawson & Lawson, 2013). This idea was supported by the findings of Lawson and Lawson in that students who felt more attached to people at their school had a greater motivation to engage in academic tasks than students who did not feel attached to people at their school.

Behavioral engagement was centered on the idea of participation (Mahatmya et al., 2012). Fredricks et al., (2004) stated that behavioral engagement was typically defined in three ways: positive conduct such as following school rules and social mores, involvement with learning activities in the classroom, and participation in school activities beyond the classroom. Behavioral engagement consisted of observable actions that entailed doing things such as paying attention, participating in classroom activities, questioning, and working on tasks (Wang et al, 2014). Fredricks et al., (2004) identified observable behaviors that included participation in activities, involvement with academic, social, and extracurricular activities, and staying on task as forms of behavioral engagement. Well-managed classrooms with expected processes and procedures were associated with higher time on task and less disruptive behavior, which were indicators of behavioral engagement (Fredricks et al., 2004). Wonglorsaichon et al., (2014) defined behavioral engagement as behaviors related to the schooling process such as completing assignments, doing as instructed, and adhering to school rules. Klem and Connell (2004)

defined it as duration of time spent on work, staying on task, and willingness to initiate action when required. Reschly and Christenson (2012) noted that behavioral engagement is sometimes split into academic engagement and behavioral engagement by researchers, with academic engagement reflecting time on task, and behavioral engagement reflecting participation.

Cognitive engagement pertained to psychological investment and willingness to put in effort (Lawson & Lawson, 2013; Mahatmya et al., 2012). Cognitive engagement entailed mental effort rather than physical and included thinking about thinking and ideas, thinking about how to solve problems, concentration (Skinner & Pitzer, 2012; Wang et al., 2014), investment in work and willingness to put in the thought required to complete assignments, being strategic or self-regulating (Fredricks et al., 2004; Lawson & Lawson, 2013; Wonglorsaichon et al., 2014) and included self-thought styles and an understanding of why they are doing what they are doing and how it is relevant to themselves (Klem & Connell, 2004).

*Consensus in findings***.** Multiple researchers found there was a significant correlation between classroom engagement and achievement (Duffield et al., 2013; Fredricks et al., 2004; Klem & Connell, 2004; Sever et al., 2014; Wonglorsaichon et al., 2014). Higher levels of engagement have led to lower chances of exhibiting disruptive behaviors (Klem & Connell, 2004) and lower rates of absenteeism (Reyes et al., 2012). Regardless of how engagement was defined, engagement was consistently linked to achievement (Conner & Pope, 2013; Skinner & Pitzer, 2012) and behavior irrespective of SES (Klem & Connell, 2004; Skinner & Pitzer, 2012). Fredricks et al. went on to add that engagement levels were found to be higher in classrooms that had supportive caring

teachers, who provided challenging and novel tasks, provided student choice, and had classroom structure.

Then consensus of longitudinal research was that student engagement decreased with progression up through high school and decreased equally for both males and females (Conner & Pope, 2013; Klem & Connell, 2004; Marks, 2000; Skinner & Pitzer, 2012). However, Conner and Pope found that levels of engagement were relatively stable up until tenth grade, at which point engagement started to decline for both males and females. From a SSPM perspective, lower engagement levels at higher grade levels was a result of a poor person-environment fit (Connell & Wellborn, 1991). Class sizes are larger (Smart, 2014) and there are lower levels of student perceived teacher warmth (Hughes et al., 2012), which based on the SSPM, result in lower satisfaction of psychological needs and levels of engagement (Klem & Connell; Skinner et al., 2008). Lower levels of teacher support at higher grade levels resulted in lower levels of engagement (Skinner et al.). Reddy et al. (2003) found levels of teacher support decreased as students progressed in age for both males and females. No study was identified that indicated males were more engaged than females at any age group (Conner & Pope; Marks).

While there is a general consensus of a positive correlation between engagement and achievement, Fredricks et al., noted varying degrees of correlation and effect in different levels of schooling, which was possibly due to use of different instruments to measure engagement, different types of students, and different measures of student outcomes. Behavioral engagement was found more likely to have higher associations with teacher grades and assessments of basic skills, while cognitive engagement was

found more likely to have higher associations with assessment that required a deeper understanding of material (Fredricks et al., 2004).

According to Fredricks et al., 2004, caution must be taken when using teachers' classroom scores of student work as a measure of outcome, as teacher perception of students, their actions, and their abilities has been shown to influence teacher assigned grades. An example of how teacher assigned grades might be influenced can be seen when comparing two students, one who is constantly causing problems but is doing the work, and the other who causes no problems and does no work (Fredricks et al., 2004). Fredricks et al., indicated the child who causes no problems is looked at in a better light by the teacher and may receive higher grades from the teacher.

In an exploratory study of 25 teachers and 9 students age 6-9 in Australia during the 2011 school year, Fried and Konzo, (2013) determined that teachers were able to commonly identify behavioral and emotional engagement in students; however, teachers struggled with cognitive engagement as the researchers noted many of the teacher assigned tasks were mismatched to student abilities and were either too difficult or too easy for students.

Ethnicity and engagement did not have direct simple linear relationships similar to age and gender (Bingham & Okagaki, 2012). There was no consensus on how engagement impacted achievement by race because there were interaction effects that depended on grade level and SES (Marks, 2000). Marks did find at the high school level, minority students were engaged more than white students. Conner and Pope (2013), however, found no differences between racial groups at the middle and high school levels. There were many factors pertaining to self-identity, culture, family support,

teacher support, school makeup, and teacher race when trying to generalize engagement levels by race (Bingham & Okagaki, 2012). Low SES students consistently showed lower levels of engagement as compared to their counterparts (Marks, 2000).

***Issues*.** Fredricks et al. (2004), in their analysis of the literature on engagement, identified many variables that were attributed to engagement and that reflect overlap with other constructs due to the way the construct is defined. Characteristics of behavioral engagement parallel previous research on student conduct and on-task behaviors. Findings from previous research on student attitudes, interests, and values were related to emotional engagement, while cognitive engagement was similar to motivational goals and self-regulated learning (Fredricks et al.). Parts of some motivation measures have engagement terminology such as self-regulation (Reschly & Christenson, 2012). Motivation and engagement are sometimes used interchangeably by researchers but are distinctly different as motivation reflects underlying energy and intention, while engagement reflects action and doing (Lawson & Lawson, 2013; Skinner & Pitzer, 2012; Reschly & Christenson, 2012). Engagement is a result of motivation (Wonglorsaichon et al., 2014), and motivation research typically includes an action component that shares characteristics with engagement which Skinner and Pitzer (2012) alluded to in stating "Engagement refers to energized, directed, and sustained action, or the observable qualities of students' actual interactions with academic tasks" (p. 24). Motivation is not observable as it is an internal private process that is an antecedent to engagement, which is observable (Reeve, 2012). Reschly and Christenson clarified that it is generally accepted that motivation and engagement are linked and influenced by context and are unique to individuals. Children who have a high level of motivation early in schooling

maintain engagement and vice versa (Skinner et al., 2008). In past studies, motivation and achievement have been highly correlated such that students who exhibit more motivation achieve at higher levels (Smart, 2014).

Fredricks et al., (2004) also identified qualitative differences in the level of engagement from low to high: Behavioral engagement can be identified as students doing the minimum of what is required or going beyond what is required and participating in activities outside of class; emotional engagement varies from a student who likes school to holding deep ties with those within it; and cognitive engagement varies from just memorizing facts to deep thought and thinking about thinking. According to Fredricks et al., “These qualitative differences within each dimension suggest that engagement can vary in intensity and duration; it can be short term and situation specific or long term and stable” (p.61).

Engagement can be measured at the school or classroom level and depends on the wording of the selected survey instrument and/or method of measurement. Many instruments mix school and classroom level items, leading to false measures of engagement, which can lead to misinterpretation of findings from an intervention (Fredricks et al., 2004; Wang et al., 2014). School level engagement and classroom level engagement are distinctly different and should be specified when presenting results as this has caused confusion in prior research (Fredricks et al., 2004). Classroom level inventories should be used to measure the effects of an engagement intervention on some variable such as achievement, to provide feedback to students in a specific classroom about their engagement levels, to determine the impact of an intervention on improving

engagement, and to get a better understanding of the impact engagement had on student learning (Wang et al., 2014).

*Study findings***.** Most studies focused on the impact of one type of engagement versus achievement, but could have used an aggregate of the multidimensional variable or the three separate measurements of engagement (Fredricks et al., 2004). According to Fredricks et al.,

> Robust bodies of work address each of the components separately, but considering engagement as a multidimensional construct argues for examining antecedents and consequences of behavior, emotion and cognition simultaneously and dynamically, to test for additive or interactive effects (p.61).

The dimensions were frequently studied independently of each other, but have been found to heavily influence each other (Fredricks et al.; Wang et al., 2014). Interventions need to focus on three types of engagement because emotion is the driver of behavioral and cognitive engagement (Skinner & Pitzer, 2012), which was also reported by Skinner et al., (2008) in stating that the dimensions of engagement were inextricably linked.

Wang et al., (2014) found in their study of 3,295 students in grades 4 through 12 across multiple subjects using the classroom engagement inventory (CEI), that there was support for the four types of engagement, with the fourth being disengagement. Their work supported the idea that compliance with school rules and norms was different from cognitive, behavioral, and emotional engagement (Wang et al.). Wang et al., also found that cognitive and behavioral engagement were two distinct dimensions of engagement since it was possible for students to do the work without thinking about it; they were just going through the motions. It is possible for a student to be doing a task to appease others, but not be interested in or enjoying what he or she is doing. Overall, their research supports the claim that higher levels of engagement lead to higher student

outcomes, and all dimensions of the CEI were positively correlated to student report card grades.

When looking at groups of students, Wang et al., (2014) determined low SES students were less cognitively and behaviorally engaged as compared to high SES students. Reschly and Christenson (2012) noted in their review that at risk students who were successful in school had significantly higher levels of engagement, as compared to those who were not successful. Girls had higher levels of emotional and behavioral engagement than boys and also exhibited less disengagement (Wang et al., 2014). Sever et al., (2014) in their research of 705 ninth to twelfth grade students in Ankara using the CEI, found there was no difference between males and females for emotional engagement; however, females reported higher levels of behavioral and cognitive engagement. They also found that students who reported higher levels of achievement also reported higher levels of engagement. In contrast, those who reported themselves as less successful were twenty times more likely to be disengaged as measured using the CEI (Sever et al., 2014).

Wonglorsaichon et al., (2014) in their research of 2,344 students using a self-report inventory of engagement, found that students ranked their level of emotional engagement highest, followed by cognitive engagement and behavioral engagement. Using structural equation modeling (SEM), their results fit the SSPM in that context influenced engagement, which then influenced achievement.

In a sample of 420 students from third to eighth grade, Klem and Connell (2004) found teacher support (caring, structured learning environment, with high expectations, rules and consequences) was associated with student engagement. Higher levels of

engagement led to better attendance and achievement. Elementary and middle school students who reported higher levels of engagement were 44% and 75% more likely to show higher levels of achievement respectively. Lower levels of teacher support, as reported by middle school students, made it more likely students were disengaged by 68%, while students who reported higher levels of support were three times more likely to be engaged (Klem & Connell).

Conner and Pope (2013) defined the construct of engagement to include the three dimensions of cognitive, behavioral, and emotional engagement, and that engagement was measured on a continuum. A sample of 6,294 students from fifteen high performing schools from middle school to high school were surveyed on variables including teacher support, engagement, and achievement. Results of the survey indicated that behavioral engagement was self-reported highest by students, followed by cognitive and emotional engagement respectively. Similar to findings in other research, females were more engaged than males across all grade levels. Unlike other research, Conner and Pope (2013) found that levels of engagement were relatively stable up until tenth grade, at which point they started to decline for both males and females; however, this study was only included high performing schools. Patterns in levels of engagement led Conner and Pope to determine that three types of students were readily apparent and included being reluctantly engaged, busily engaged, and fully engaged, with fully engaged students showing higher levels of the three dimensions of engagement. Students that were categorized as fully engaged had significantly higher GPA’s than other students, and also reported greater TSR. TSR was also found to be correlated to all three engagement dimensions.

Using self-report surveys from 805 students and 53 teachers, all predominantly white, from fourth to seventh grade and at the beginning and end of the year, Skinner et al., (2008) confirmed the findings of prior research. Older students were less engaged and showed more signs of disaffection as compared to younger students. Students reported a decline in teacher support while progressing through schooling. Students that reported greater levels of teacher support had higher levels of behavioral and affective engagement. Again, females were found to be more engaged than males, but levels of engagement dropped off similarly to males. Students who were emotionally engaged at the beginning of the year had higher levels of behavioral engagement at the end of the year. Autonomy was a strong predictor of change in both forms of engagement, but more so for emotional engagement. Engagement levels were shaped by teacher support and student self-perceptions. Behavioral and emotional engagement were linked to each other with emotional engagement fueling behavioral engagement (Skinner et al.).

Marks (2000) used hierarchical linear modeling (HLM) with nested classrooms within nested schools across a nationally representative sample of elementary, middle, and high schools of 3,669 students in math and social studies classrooms. She found that engagement levels declined with age and that females were more engaged than males. In this study, there was no difference in engagement across race. Marks also determined that classroom social support and authentic work experiences had a positive influence on engagement.

TSR, engagement, and achievement. Teacher-student support was shown to influence the three types of engagement, with higher levels of engagement resulting from higher levels of support from teachers and peers (Fredricks et al., 2004). In middle

school, teacher caring had lasting effects on student engagement when controlling for previous academic performance (Furrer & Skinner, 2003). Students that reported greater levels of teacher support had higher levels of behavioral and affective engagement (Skinner et al., 2008). Teacher caring and support was positively associated with participation, on-task behaviors, and less acting out by students, which is part of behavioral engagement, which in turn, influenced the student relationship with teachers (Fredricks et al., 2004). Environments in which students were emotionally supported had greater levels of engagement, even after controlling for achievement with Reyes et al. (2012) also finding students reported greater interest and enjoyment of class and achieved higher scores on standardized assessments. Classrooms in which students felt greater cohesiveness, satisfaction, goal direction, less disorganization, and less friction had greater student achievement (Henderson, 1995). Emotional engagement was higher when students' need for relatedness was more satisfied (Fredricks et al., 2004).

Reyes et al., (2012) found classroom emotional climate (CEC) was linked to achievement both directly, as a proximal process to achievement, and indirectly, as mediated by engagement with higher levels of CEC showing higher levels of achievement. For every 1 unit increase in CEC, there was a 3.83 point increase in achievement, which equates to half a letter grade. Student engagement mediated the effect of CEC on achievement. Higher student engagement was associated with higher achievement, with a 1 unit increase in engagement equating to a 1.74 point increase in achievement. Classrooms high in emotional support were thought to support connectedness and belongingness to the classroom, enjoyment, and respect in the classroom, which were related to a student's underlying psychological needs of

autonomy, competence, and relatedness. Classrooms high in emotional climate provided a safe and enjoyable place to be, which resulted in students becoming more engaged in learning and scoring higher. The finding that CEC was significantly related to engagement supported SDT in that engagement and achievement were not solely the responsibility/fault of the student, but also the classroom context (Reyes et al.).

Reyes et al., (2012) found that instructional support, classroom organization, teacher demographics, and teacher experience had no statistically significant impact on engagement and achievement. Reyes surmised the finding of insignificance of classroom organization and instructional support as compared to engagement and achievement was possibly due to use of the CLASS external observer tool, which may not have captured classroom organization and instructional support from an outsider's observation of three classes. This issue might be resolved by having a greater number of observations or observations from a person within the school environment. Another limitation included the possibility of high shared variance between teachers that had high CEC classrooms and student scores. It is possible that teachers scored students higher because they were more emotionally connected to those students and took that into account when doing grading. To counteract the possibility, Reyes et al. recommended using standardized test scores in future research.

Reddy et al., (2003) investigated how teacher support influenced teacher-student relationship quality, as student perceived ratings of support have been shown to be more related to student outcomes than actual help received. Reddy et al. found, consistent with other research, levels of teacher support decreased as students progressed in age for both males and females. Females, however, reported higher initial levels of teacher support,

possibly due to being more attuned to interpersonal cues in relationships with teachers. This research was limited since it was not teacher and classroom specific with measurement instruments, but rather captured school level relationship quality information. Students may have been influenced by teachers when filling out the survey instrument as there were also teachers in the room during survey administration.

Student growth percentiles as an outcome. In the research previously cited, various measures of student outcomes included attendance, engagement, and achievement in the form of GPAs, class averages, teacher test scores, and standardized status scores. Student growth as determined by student growth percentiles, a recently adopted measure, has not been utilized as an outcome in prior research. While no peer reviewed literature was available at the time of this writing, there were three dissertations that studied the relationship of classroom variables and their influence on student growth as determined by student growth percentiles and one on student growth as determined by gain scores.

Cervoni (2014) investigated the relationship between factors found to have a large influence on student achievement that were encouraged by the state of New York such as differentiated instruction, group work, encouraging student engagement, use of formative assessments, years of teaching experience, educational levels and the practices' impact on student growth percentiles. Cervoni was surprised by the results and stated, "Stunningly the results suggest that none of the practices reported in the study appeared to have any effect at all on student growth percentile scores" (p. 91). Aggregated growth scores for teachers were well below the 50th percentile, yet their self-report surveys indicated they had implemented many of the practices listed above. The findings, however, may be

limited as the researcher had a small sample size and no details about the school's prior achievement levels, limiting the applicability of the findings.

Craig (2011) investigated report card format and the impact on student growth percentiles in elementary schools in Massachusetts. The hypothesis was that a standards-based report card provided effective feedback and promoted self-efficacy and motivation resulting in higher growth. Craig, in a causal comparative study of 103 elementary schools, found using standards-based report cards had no impact on student growth percentiles in math. There were, however, some indications in the research that standards-based report cards did positively influence low SES and special education students' student growth percentiles, but the findings were not statistically significant.

LeGeros (2013) focused on the relationship between student growth percentiles and elementary math teacher licensure exams, as a measure of teachers' content knowledge, in the state of Massachusetts with a sample of 130 teachers and 2640 corresponding students in grades four and five. Three natural groups of teachers existed in Massachusetts and included teachers that fully passed the MTEL test (score > 240), teachers that conditionally passed the MTEL test (score > 227 and < 240), and teachers that failed the MTEL test (score < 227). Students with teachers who conditionally and fully passed the MTEL had statistically significantly higher student growth percentiles than students with teachers who failed the MTEL test. Passing the MTEL state licensure exam showed a teacher had detailed content knowledge, and resulting instruction influenced student growth in the classroom.

Using Stronge's characteristics of effective teachers, which included caring, fairness, respect, interactions with students, enthusiasm, reflective practice, and

motivation for learning and administrator evaluations of teachers based on Stronge's characteristics, Simmons (2006) evaluated how teachers deemed effective influenced student growth on the Idaho Standard Achievement Test from fall of 2004 to spring 2005. Student growth was measured as a gain score by subtracting fall results from spring results. Level of education and teaching experience had no impact on student growth in math or reading. Teachers evaluated as effective, according to Stronge's "teacher as a person" traits by administrators, had no statistically significant impact on student growth in math or reading. Similar to Craig (2011), Simmons did find a statistically significant but weak positive association between low SES students and math growth scores.

Evaluation and Accountability

School district, teacher, and student performance can be measured and evaluated using status models and/or growth models (Batista, 2014). The goal of NCLB was to ensure proficiency of all students by 2016 in reading and math based on status scores (Nichols et al., 2005; Ladd & Lauen, 2010). Status models rate student performance based on a student's current status (i.e., achievement level). According to Betebenner (2008), "Status models are unconditional achievement models, examining student performance at a point in time with no conditioning variables" (p. 2). Accountability systems constructed under NCLB according to adequate yearly progress (AYP) were based on student achievement measures of reaching terminal objectives yearly. Joshua et al., (2006) summarized NCLB when they stated, "Testing or measurement has been a process of gathering quantitative estimates of the amount of knowledge, skills, traits or characteristics possessed or acquired by learners in the school system" (p.1). Standardized assessment status scores were indicators used under the No Child Left

Behind Act of 2001 for school district accountability (Nichols et al.; Betebenner) with obtained data then used to make decisions on administration, instruction, and learning (Joshua et al.).

The teacher is the most important determinant of student learning in the classroom; therefore, test scores are a measure of teacher effectiveness (Darling-Hammond, 2015; Joshua et al., 2006). According to Haertel (2013), 9% to 13% of variance in student achievement was determined by a student's teacher with 60% of variance in achievement accounted for by factors outside of the school's' control. In the past 10 years, accountability has shifted from a focus on schools' effectiveness to a focus on teacher effectiveness. This shift has brought with it a move from accountability using status scores at the school level to accountability using growth scores at the teacher level. Race to the Top (RttT) funds have given financial incentives to states to develop accountability systems that link student growth to teachers and teacher evaluation systems (Collins & Amrein-Beardsley, 2014; USDOE, 2009). In November 2005, the U.S. Secretary of Education started the Growth Model Pilot Program (GMPP), which allowed states to use growth model results instead of status measures to meet NCLB mandates (Betebenner, 2008).

Status Model

Status models compare student performance to targets set according to Federal adequate yearly progress (AYP) (Thurlow et al., 2010) with annual targets being increased every year for the percentage of students meeting proficiency (Ladd & Lauen, 2010). Accountability systems based on AYP relied on evaluating annual snapshots of student achievement to judge school quality (Betebenner, 2011; Doran, 2003). Status

scores at a single point in time on standardized assessments provided snapshots of students' ability (Doran; Ladd & Lauen). Status models evaluated student's or a cohort of students' achievement at one point in time (Castellano & Ho, 2013) and compared it to an established target as measured by percentage of students meeting or exceeding set goals. Status models compared a cohort's progress from year-to-year of the same class and grade to determine if the cohort improved or not. For example, this year's 3rd grade math scores will be compared to next year's 3rd grade math scores (Thurlow et al.). A look at longitudinal data showed whether schools had a greater percentage of students proficient or not proficient on assessments. There must also have been a decrease in differences in subgroups for race, SES, and SWD (Ladd & Lauen). Fewer students achieving proficiency was judged as the school being less effective and underperforming (Goldschmidt, Roschewski, Choi, Auty, Hebbler, & Williams, 2005) even though a student may have started the school year lacking prerequisite skills to be successful.

Status models did not and do not adjust for or take into account any preconditions of students. With a social efficiency mindset, status models efficiently identified the percentage of students meeting the terminal objectives (Thurlow et al., 2010) regardless of prior academic achievement. Status models made evaluation easy, as data was readily available from high-stakes testing results (Batista, 2014). Student score reports informed students, parents, teachers, and the public that objectives were or were not met (Doran, 2003). Descriptive statistics were used to determine and display current and past achievement level of groups of students and were easily displayed and interpreted by the public (Thurlow et al.) to whom schools are accountable (Schiro, 2013). Betenbenner (2008) stated,

> The output from such models, within assessment systems found in all states, were usually a simple qualification of achievement for each student based upon the state's performance standards. As the basis for an accountability system with rigorous achievement standards, such models were extremely demanding, requiring without condition, an acceptable level of achievement from all students (p. 2),

identifying that status models had no regard for student abilities and school effectiveness.

Use of and interpretation of status scores presented many problems for educators. This unconditional evaluation model of schools was problematic because use of a single indicator to rate students, teachers, and schools was not a good indicator of what schools were really doing for their students (Nichols et al., 2005). Doran (2003) noted descriptive statistics were used to analyze status scores and were not informative since aggregated scores could not provide information on how to improve instruction. Doran further went on to state that score reports on students informed the teacher whether objectives were met or not, but did not show where the student/class was excelling or lacking and did not allow for remediation without starting at the beginning. Status scores did not show individual student improvements, just cohort improvements (Thurlow et al., 2010).

Cut scores for proficiency were set arbitrarily by states and typically had very wide ranges, which made it difficult to conclude the true level of a student's academic ability (Batista, 2014; Doran, 2003; Ladd & Lauen, 2010). It was possible that a school could do a poor job of educating very capable students and be rated too high, and that a school could do a good job of educating students that were not ready for the skills to be learned and were rated too low (Batista, 2014).

The pressure of using this indicator should have led to increased student achievement as teachers/schools did not want to face sanctions, but instead, sometimes

led to corrupt behaviors (Nichols et al., 2005) such as the Atlanta erasing scandal in which 178 teachers and administrators changed student answer documents to increase standardized test scores (Osunsami & Forer, 2011, July 6). An unforeseen consequence of NCLB and evaluation based on status scores has led teachers to focus on the test and test taking skills (Nichols et al.) and focus much of their attention on “bubble students” that were nearest the proficiency cut score (Doran, 2003; Ladd & Lauen, 2010; Thurlow et al., 2010) in order to show yearly improvements..

Status scores did not show the effects of schooling imparted on students (Thurlow et al., 2010). Snapshots of student achievement contained measurements of both school and non-school effects (Doran, 2003). Many factors that hindered/encouraged student achievement were outside the control of the school and classroom teacher, which made interpreting test scores difficult (Joshua et al., 2006). Standardized assessments in specific grade levels did not measure skills just from that grade level, but also all prior grade levels, making it possible that student learning came from prior teachers (Doran). Accountability based on status scores had an unfair expectation that schools can make up for non-school factors such as a student’s or school’s socioeconomic status (SES) (Ladd & Lauen, 2010). Stratification by SES in the U.S. has impacted students’ test scores and unfairly targeted schools with a greater percentage of low SES students (Darling-Hammond, 2015; Doran; Ladd & Lauen). Haertel (2013) stated,

> School climate and resources, teacher peer support, and of course, the additional instructional support and encouragement students receive both out of school and from other school staff all make the task of teaching much easier for teachers in some schools and harder in others (p.11).

Student growth was neglected, whether or not a student met proficiency levels. A student may not have met the cut score but may have shown a great deal of learning in the class,

which was disregarded as unacceptable and worked against schools that served low SES students (Doran, 2003; Thurlow et al., 2010), thus this made it nearly impossible for some teachers to have what was considered high achieving students based on status scores (Haertel, 2013).

Growth Model

Shifting away from status models, RttT (Collins & Amrein-Beardsley, 2014) along with the Growth Model Pilot Program (Betebenner, 2008) have fostered development and implementation of growth-based accountability systems throughout the United States. Forty states are or will be using growth models in one form or another as components of teacher evaluations, with thirty of those having state legislation that requires its use to measure part of teacher effectiveness (Collins & Amrein-Beardsley, 2014; Darling-Hammond, 2015). Using student growth data to inform teacher evaluations has become an integral part of education system reform under the *2009 American Recovery and Reinvestment Act* (Ryser & Rambo-Hernandez, 2014).

The primary purpose of the growth model is to provide insight into student learning and be able to attribute student learning to a teacher/principal/school (Betebenner, 2008; Doran, 2003). Schools should be held accountable for student learning they can control within the context of the school year (Ladd & Lauen, 2010), and growth models do a better job disentangling the teacher contribution portion of student growth (Guarino, Reckase, Stacy, & Wooldridge, 2014). Growth models track and describe academic performance and measure the achievement of individual students over two or more points in time (Castellano & Ho, 2013; Doran; Ladd & Lauen; Thurlow et al., 2010) compared to status scores that show achievement at only one point in time

(Betebenner, 2011), making growth models better indicators of teacher performance (Briggs, Dadey, & Kizil, 2014b). With the inclusion of growth model data, more information will be gained about student learning and teacher impact than using status scores alone (O'Malley, Murphy, McClarty, Murphy, & McBride, 2011). While there are reasons as to why a student may not have an acceptable status score beyond a teacher's control which growth models attempt to disentangle (Batista, 2014; Doran; Haertel, 2013), there is no reason a student should not grow from one year to the next, even if there is only little growth (O'Malley et al., 2011). It is important to know that even using growth models to measure accountability, school districts will not meet 100% universal proficiency under the NCLB mandate (Betebenner), and the achievement gap will still exist between white and non-white students, low SES and high SES students, and students with disabilities and students without disabilities (Bingham & Okagaki, 2012).

Growth models are ideally suited for educational purposes because they are philosophically aligned with educators' viewpoint that they teach students and get them to grow (Thurlow et al., 2010). Thurlow et al., further identified that growth models highlight the missing point of the status model, in that even though the student may not have reached the set goal, he or she grew towards the set goal while also accounting for different starting points of different students. Using growth model data and longitudinal data, the teacher can now make decisions on how to aid his or her students (Doran, 2003).

While growth models elucidate greater information about students, teachers, and schools, there are issues that make using growth model information limited. Some growth models are complex, require complex statistical methods and models, and are not easily explained to educators and the public (Thurlow et al., 2010). Any growth measure

must meet minimum statistical assumptions, and violation of any of the assumptions reduces the validity of findings and makes interpretation of accuracy difficult (Haertel, 2013). Most growth models are criticized because there is no requirement to meet set standards (Ladd & Lauen, 2010) since the key metric is that a student grow. Low status scores may become acceptable so long as growth is acceptable (Thurlow et al.), which Betebenner (2008) noted in referring to the fact that districts will not meet 100% universal proficiency under the NCLB mandate. Finally, growth models typically require use of standardized assessments. Roughly 69% of the teacher population teaches subjects that do not have standardized assessments and, therefore, cannot have growth model data for their classrooms (Prince, Schuermann, Guthrie, Witham, Milanowski & Thorn, 2009).

Growth models are as reliable as conditions permit and only provide valid data when students are assigned randomly to districts, schools, and classes, and individual teachers are the only contributors to student learning, all of which are nearly impossible to achieve (Darling-Hammond, 2015). Darling-Hammond points out that assessments are built for grade level standards and are biased against low and high ability students. She also noted that society in the U.S. is stratified by race and socioeconomic status, allowing some schools to have more resources than others. Haertel, (2013) stated that “No statistical manipulation can assure fair comparisons of teachers working in very different schools, with very different students, under very different conditions” (p.24). Many prior studies on a variety of growth models indicate unstable estimates of year-to-year teacher contributions to student growth (Hammond-Darling) with correlations ranging from .2 to .5 (Haertel).

Teacher Accountability

Schools and teachers should be held accountable for learning they can control within the context of the school year (Ladd & Lauen, 2010) with student learning as indicative of teacher effectiveness (Darling-Hammond, 2015). Ehlert, Koedel, Parsons and Podgursky, (2013) have shown that teachers have a dramatic impact on student growth. With the shift of accountability moving to teachers and requirements of RttT, teacher and principal evaluations are being tied to student growth as part of accountability systems and counts as much as 50% towards a teacher's evaluation (Collins & Amrein-Beardsley, 2014).

Student growth is a better performance metric of school and teacher quality, with effective schools and teachers bringing about student growth and non-effective schools and teachers not bringing about student growth (Betebenner, 2011). Growth models measure progress by tracking achievement scores of students rather than cohorts to determine if individual students are making progress. Student growth can then be compared to others and to statewide or local targets (Betebenner, 2008). Aggregated growth data can be used to describe group growth of classes, subjects, schools and/or districts (Bylsma, 2014; Castellano & Ho, 2013). Along with using student growth scores, many studies have indicated that observations and feedback allow teachers to become more effective in developing ways to teach and assess their students (Darling-Hammond, 2015).

As a requirement to RttT, student growth is now a part of the Teacher Keys Effectiveness System (TKES) in Georgia (USDOE, 2009), with student growth accounting for fifty percent (GaDOE-CIA, 2014b). According to Briggs et al., (2014),

> The inference to be made is that a student who has performed better/worse than comparable peers has demonstrated more/less academic growth. If the average student in a teacher's class tends to demonstrate performance on subject-specific tests that is above/below that of peers with similar prior academic achievement, it suggests that the quality of teaching the student experienced may have also been above/below average. This is formally quantified for each teacher in the TKES by taking the mean of SGPs (a "MeanGP") across students (p. 2).

Growth Models

There are a variety of growth models used in the U.S., with different assumptions and different ways of determining student growth (Goldhaber et al., 2014), and they typically fall within the two categories of value added or growth models (Collins & Amrein-Beardsley, 2014; Guarino et al., 2014; Thurlow et al., 2010), both with the purpose to strip away factors outside of a teacher's control and to determine how much the teacher influenced student achievement (Haertel, 2013). People without an extensive background in either model may use value added and growth model synonymously to mean the same thing, but they are in fact quite different in what they describe and how they are determined.

Both value added and growth models use prior-year achievement scores as a covariate of current year growth and have high associations with each other (McCaffrey & Castellano, 2014). In a review of 7 studies by McCaffrey and Castellano, which included different states, students, and grade levels, value added and growth models had correlations of .77 to .93 for teachers and .69 to .99 for schools, which were stronger with inclusion of a greater number of prior years of student test data. McCaffrey and Castellano found that a majority of teacher ratings would not differ drastically using one model or another; however, some teacher and school ratings were notably different,

consistent with the findings of Guarino et al., (2014) in their study of fifth and sixth grade math students in a single school district.

Value added models (VAM) and student growth percentiles both had reliability issues when bias was not accounted for in the form of non-random classroom assignment of students (Guarino et al., 2014). When students were not randomly assigned to classrooms, Guarino et al., found there were lower correlations between VAM's and SGPs along with faulty conclusions of teacher effectiveness based on SGPs.

Value added models use and control for preexisting differences and characteristics of school and non-school related factors among students when determining the impact of the teacher and school on student growth and attempt to isolate the cause of student growth (Betebenner, 2011; O'Malley et al., 2011). Factors included in VAM's include prior academic achievement, sex, race, English language status, student disabilities, and socioeconomic status (McCaffrey & Castellano, 2014). Growth models do not account for preexisting conditions, are descriptive (Buzick & Laitusis, 2010; Goldhaber et al., 2014), and do not try to isolate the cause of student growth (Bylsma, 2014; McCaffrey & Castellano; O'Malley et al.).

Value added models focus on the teacher and school while growth models focus on the student. Value added supports causal inferences of who impacted student learning while growth models are descriptive and describe growth compared to others (Buzick & Laitusis, 2010). A value added estimate is the difference between actual growth and expected growth (Betebenner, 2011) with the possibility to have a negative score when the student grows, but not as much as expected (Goldschmidt et al., 2005).

Student growth percentiles. Currently, student growth percentiles (SGP) are the most commonly used growth model in the United States being used or piloted by 12 states (Collins & Amrein-Beardsley, 2014) including the state of Georgia (GaDOE-CIA, 2014b). Georgia has implemented student growth percentiles (SGP), similar to Colorado, Massachusetts, Indiana, Wisconsin, and Hawaii (Buzick & Laitusis, 2010), which will provide the additional perspective of growth data on top of status data. SGP data will provide more detailed information on student learning, improve both a teacher's teaching and a student's learning, inform teachers and administrators of educator effectiveness within TKES and LKES, and be used as multiple indicators as part of College and Career Readiness Performance Index (CCRPI), which is the state accountability system (GaDOE-CIA, 2014b). The Georgia Student Longitudinal Data System, which houses all student information at the state level, will present growth scores calculated using median SGPs, while mean SGPs will be used in student growth calculations for measures of teacher effectiveness in the TKES process (GaDOE-CIA, 2014b).

Betebenner (2008) introduced student growth percentiles as a normative approach for describing student growth (Castellano & Ho, 2013; Wyse & Dong, 2014) and not to determine the causal impact of the teacher (Guarino et al., 2014). SGPs were intended to have a correlational, not causal, relationship with teacher effectiveness. If a teacher class had high growth, it might have been attributable to the teacher, but it might also have been attributed to other factors (Briggs et al., 2014b).

SGPs examine current student status as compared to prior student status of “academic peers” and places measures of growth on a more level playing field as students are compared to similarly achieving students based on past performance (McCaffrey &

Castellano, 2014). Location on current assessment with respect to “academic peers” is expressed as a percentile rank. (Castellano & Ho, 2013; Buzick & Laitusis, 2010; McCaffrey & Castellano, 2014). SGPs do not show an exact amount of growth from one year to the next and only give a relative standing judgment as compared to academic peers (Betebenner, 2011), as Georgia standardized assessments are not vertically aligned (GaDOE-CIA, 2013). Growth is conditional based on the prior scores of a student as compared to their “academic peers” (Briggs et al., 2014b; Buzick & Laitusis; Ehlert et al., 2013; Wyse & Dong, 2014).

In order for students to have an SGP calculated, students must have a prior- and current-year test score in the same subject such as, from seventh grade math last year and eighth grade math this year (Briggs, Dadey, & Kizil, 2014a). Similar to other types of value added and growth models, SGPs can be based on many years of prior growth data (Wyse & Dong, 2014), but typically have a maximum of two years of prior test scores (GaDOE-CIA, 2014b; USDOE, 2011a). SGPs are whole numbers ranging from 1-99 (Betebenner, 2008) with low and high performing students capable of attaining any score in that range (Castellano & Ho, 2013). Georgia has set four levels of student growth to distinguish between low and high growth: An SGP of 1 to 29 is categorized as level I, 30-40 as level II, 41 to 65 as level III, and 66-99 as level IV (GaDOE-OSI, 2014b). A growth score of 45 is interpreted as a student scored better than 45% of his/her academic peers based on the prior year test score. The SGP of 45 cannot be interpreted as absolute growth and is considered normative as it is in comparison to academic peers (Castellano & Ho, 2013) and has no bearing on a student's level of achievement on a standardized assessment. SGPs do not provide information about how a student is progressing as

compared to a set bar; therefore, an SGP of 45 for one student may not have the same meaning for a different student with an SGP of 45 (Betebenner, 2011; Bylsma, 2014). It is entirely possible for one student to have lower growth, not because he or she learned less, but because the other student scored much higher.

Issues with student growth percentiles. All growth models are based on statistical methods and models; therefore, there are assumptions and limitations of SGPs to be aware of when utilizing the statistical procedure and interpreting results. Use of SGPs as a metric of teacher performance are not without issues because SGP group metrics designed for high stakes decisions in education may be subject to corruption, inflation, and gaming (Castellano, & Ho, 2013). Normative growth of SGPs have not been interpreted correctly by many due to non-vertical scale of state standardized assessment (O'Malley et al., 2011); however, growth percentiles are easily explained to the lay person once scaling of assessments is relayed (Betebenner, 2008; Bylsma, 2014). SGPs are calculated using quantile regression, not linear regression, and therefore are not predictive (Wyse & Dong, 2014).

Another concern is that students that do not have prior years' assessment scores will not receive a SGP and, consequently, will not be included in the aggregated SGP at the classroom or subgroup level (GaDOE-CIA, 2014b), which could pose a problem for schools with a more transient population. An additional issue lies in the fact that SGPs cannot be interpreted as causal like value added models (Betebenner, 2009). For example, teachers at school A having had higher aggregated growth scores than teachers at school B does not mean that teachers are better at school A, as school demographics and SES could have been different (Castellano & Ho, 2013). Castellano and Ho further

stated, for growth models to support value added claims, several years of test data for the same educator with a large number of students are needed.

Unlike VAMs, SGPs include no controls for student characteristics found to influence achievement such as race, socioeconomic status, sex, prior achievement levels or schooling environments like value added models (Bylsma, 2014; McCaffrey & Castellano, 2014; Wyse & Dong, 2014) as adjusting SGPs on such characteristics lower expectations for certain groups (Ehlert et al., 2013). Classroom contexts of the proportion of low SES students, proportion of ELL students, and proportion of SPED students is not accounted for in Georgia SGP, because this model does not attempt to adjust for these factors (Briggs et al., 2014a; Briggs et al., 2014b), which can be problematic.

Growth model data can be used to improve instruction by reinforcing positive educational practices and discouraging negative ones. Teachers with students consistently having high growth model scores can be observed to identify what factors are working in that classroom. Teachers with consistently high growth scores should be mentors and serve as models of excellence (Ehlert et al., 2013). While growth data can be used to improve teacher performance, there was no evidence that providing teachers growth scores about their students would increase their ability to understand the information and use it to improve classroom skills and instruction (Collins & Amrein-Beardsley, 2014). Collins and Amrein-Beardsley identified that of the states implementing growth or value added models, no state representative was able to articulate a plan on how the data would be used to improve teacher effectiveness.

Conclusion

According to the review, among other variables, teacher-student relationships, basic psychological needs satisfaction, and engagement influence student achievement in schools. The seminal work of Hattie (2009), Roorda et al., (2011), and Cornelius-White (2007) identified teachers as the most important influence of student achievement in the school system, with 70% of the variability in student achievement due to student perception of interpersonal teacher behavior (Wubbels & Levy, 1993).

The hypothesized Self-systems Process Model proposed by Connell and Wellborn (1991), which is based on Deci and Ryan's (1985) empirically well-supported self-determination theory, illustrates how the social context influences basic psychological needs satisfaction (self) to influence student engagement (action) and, consequently, achievement (outcome). Other than Connell and Wellborn, there is no research testing the validity of the full model. Hattie (2009), Roorda et al., (2001), and Cornelius-White (2007) identified that teacher-student relationships impact achievement directly and indirectly through engagement. There is consensus in the literature that engagement, however it is defined or measured, is a proximal process to achievement, and when demonstrated by students, leads to higher levels of achievement as measured by student class averages, GPA's, standardized assessments, and teacher generated assessments (Fredricks et al., 2004; Klem & Connell, 2004; Skinner et al., 2008). There was, however, no research on how teacher-student relationships, basic psychological needs satisfaction, and/or engagement influenced student growth percentiles. There were three dissertations on variables that influence SGPs, but these did not include teacher-student relationships, psychological needs satisfaction, or engagement.

With the shift in accountability moving to teachers, how teachers are evaluated and deemed effective or not have changed drastically in the state of Georgia with student growth accounting for a significant portion of a teacher's evaluation (GaDOE-OSI, 2014b). Districts, schools, their leaders, and teachers have the responsibility to recognize the changes in teacher evaluation and measurement of student growth. The aforementioned school personnel has the responsibility to investigate and evaluate the effectiveness of classroom strategies as it pertains to student growth percentiles in Georgia because fifty percent of the teacher evaluation is now predicated on student growth. In order for teachers to improve their overall evaluations, they must improve their students' growth scores, on which there is little to no research, which justifies this study.

This research will address both the lack of literature pertaining to variables that influence student growth percentiles and to evaluate the full Self-systems Process Model. This research will build on prior findings in the research utilizing standardized assessment status scores as the dependent variable, and then comparing the results with an identical methodological setup with student growth percentiles as the dependent variable.

CHAPTER III
METHOD

Introduction

Researchers have found that teacher-student relationships (TSRs) significantly influence student classroom engagement and student achievement through satisfaction of basic psychological needs. Connell and Wellborn (1991) provided evidence supporting their hypothesized Self-Systems Process Model (SSPM), which showed a path from context to self, action, and outcome, which is based on the premise of supporting an individual's basic psychological needs. In their model, student perception of psychological needs satisfaction mediated the effect of the teacher-student relationship on student engagement. The current research will not utilize teacher perceived measures of the TSR, basic psychological needs, and engagement and will solely rely on student perception data of teachers, which has been used in research dating back to 1896 (Burniske & Melbaum, 2012). Other research has identified that student engagement mediated the effect of perception of psychological needs on student achievement as measured by GPA, standardized assessments, and end of course grades.

This quantitative study examined the extent that teacher-student relationships influenced basic psychological needs, engagement, and growth/status scores using the SSPM as a framework, with the outcome being measured using the Georgia Milestones

standardized assessment norm-referenced scores, class GPA, term 4 student average, and student growth percentiles.

The research was guided by the following three research questions:

1. To what extent does the teacher-student relationship influence satisfaction of basic psychological needs which influence engagement and, consequently, influence student growth percentiles as compared to student status scores using an identical methodological setup (Context → Self → Action → Outcome)?

2. To what extent is the effect of teacher-student relationships on student growth percentiles invariant across population subgroups? (i.e. Low socioeconomic status students versus high socioeconomic status students and White students versus non-white students)

3. To what extent does the teacher-student relationship influence level of student engagement (Context → Self → Action)?

Participants

This study took place at one medium to large rural school district in southwest Georgia at the sole middle school in the district. The total student population of the school district was 5,218 in grades Pre-K through twelve. The population was 73.4% white, 16.6% African-American, 3.4% Hispanic, 5.1% multiracial with 29.7% of the students receiving free lunch and 6.3% of the students receiving reduced-price lunch and were marked as low socioeconomic status students. The primary participant pool (n = 809) consisted of seventh and eighth grade students.

Recruitment letters and parental informed consent (see Appendix D) to participate forms were sent home with all seventh and eighth grade students along with their third

nine week report cards following IRB approval. A total of 40 parental informed consent forms were returned to student's homeroom teachers for an initial participation rate of 4.9%, which was well below the number of participants needed to conduct an SEM analysis. The school administrator informed the researcher that since no teacher "owned" or was responsible for the forms, there was no one pushing for students to get the document signed and returned to the school.

Based on the principal's recommendations, the researcher presented the research opportunity to all teachers of the school and solicited volunteers to assist in getting teams of students to return parental informed consent forms. One seventh grade English language arts, two seventh grade social studies, one eighth grade English language arts, and one eighth grade science teacher volunteered to "own" the forms and assist the researcher in getting students to return signed informed consent forms. These teachers represented five of the eight teams at the middle school level with a participant pool of $n = 504$. A second set of parental informed consent forms were sent home with these students on these five teams and a total of 218 forms were returned for an overall school response rate of 31.9%.

Instruments

All students that returned parental informed consent forms participated in the research process, which started on May 2nd, 2016, and took place over the course of five days for two weeks, to accommodate both the researcher's and school's schedule. Throughout each day the survey was administered, the researcher went to the classroom and escorted participants to the survey room. Classroom teachers were not present during survey administration, and only the researcher was present during every administration.

Prior to completing the survey, students were handed assent forms (see Appendix E) and the researcher went over the assent form by reading the information to students and answered any questions students had. Students were made aware that they did not have to participate in the research and that there were no repercussions for not participating. Only one student chose not to participate, all other students agreed to participate and signed the student assent form wherein they were informed, that if they did not feel comfortable responding to the survey or a specific question, to quit the digital survey.

Once students signed the assent document, they were instructed to open the iPad and click on the survey icon to load the instructions for completing the survey. To improve the quality of data collected, students were read standardized instructions by the researcher and completed the survey without their teacher present. They were informed of the importance of their honesty in their responses as the research may guide future practices in schools, and they were assured that the results would be strictly confidential.

Data collection was conducted through Google forms using the researcher's personal Gmail account, to which no one else had access. Student responses to the form were automatically collected and placed in a spreadsheet that was inaccessible by anyone other than the researcher. The form consisted of 5 pages, with a page for student information (4 items), two pages for teacher-student relationships (12 items each), a page for basic psychological needs (9 items), and a page for classroom engagement (13 items) and can be viewed at https://goo.gl/0Rrpul. To alleviate issues of missing responses and using the capabilities of Google forms, the digital questionnaire was setup to require responses for all questions on each page before moving on, which eliminated missing responses from the raw data file. In an attempt to get response data for all core academic

subjects in seventh and eighth grade, students were informed to complete the survey two times, once for the teacher of class they were currently in, and once for the class they would attend next period. Completing the two surveys took students about 25 minutes.

Wireless connection issues caused eight student surveys to time out and students had to restart the survey. Responses entered by students prior to the survey timing out were recorded by Google, but were missing entire section(s), which resulted in eight incomplete student surveys. These response sets had to be deleted as imputation would have been impossible due to entire sections of indicator items of latent variables missing. One student entered an invalid student ID of 5555555 and could not be matched with either grade point average information or Georgia milestone assessment results. One student did not participate in the 2015-2016 Georgia Milestones Assessment, and eleven students either transferred in from another state or private school and had no growth scores and were also removed from the dataset, leaving a total of 512 student responses.

Student Georgia Milestones assessment scale scores, norm referenced scores, growth scores, fourth term averages, and final GPAs were collected through the district's curriculum department and merged with the spreadsheet of student responses based on the student ID. Student demographic data consisting of gender, race, and socioeconomic status was also merged with survey data at this time.

Measures

Demographic information. Basic student information was collected that consisted of student ID, grade level, academic area, and teacher last name. The student ID was used to collect information about student gender, race, and socioeconomic status. The

breakdown of responses by race and socioeconomic status mirrored that of the school and district; however, there were more responses by females and seventh grade students.

Network of relationships inventory. There was no inventory available that measured the teacher-student relationship from a student perception. The STRS by Pianta (2001) purported to measure the relationship, but from a teacher point of view. The teacher-student relationship was measured using the Network of Relationships Inventory (NRI), which was developed to examine characteristics of an individual's relationships with others (Furman & Buhrmester, 2009). Furman and Buhrmester developed three different versions of the NRI, which included the Social Provisions Version (NRI-SPV), the Behavioral Systems Version (NRI-BSV) and the Relationship Qualities Version (NRI-RQV), all of which could be used to examine various relationships and the association with specific outcomes. According to the authors, the inventories are appropropriate for children 11 years and older.

In this research, the NRI-RQV was used to measure student perception of closeness and discord in their relationship with their teacher, as the NRI-RQV was developed to describe supportive and discordant qualities in relationships between children and adults. The inventory was then used to compare how individual students perceived their teacher and the impact on satisfaction of basic psychological needs, engagement, and student outcomes.

The original NRI-RQV consisted of 30 questions and used a 5-point Likert scale based on frequency ranging from 1 to 5 (Never or hardly at all to Always or extremely much) to measure two second order factors of closeness and discord in relationship quality (Buhrmester & Furman, 2008). There are ten first order factors, five assessing

positive aspects of the relationship which include companionship, disclosure, emotional support, approval, and satisfaction; and five assessing negative aspects of the relationship and include conflict, criticism, pressure, exclusion, and dominance with all scales having three items. Each of the ten scales are scored by finding the mean of the three items. The factors of closeness and discord can be determined by finding the mean of the included scales.

According to Furman and Buhrmester (2009), the inventory can be altered to specify the relationship to be measured and to eliminate unneeded scales. All questions included in the questionnaire were modified to include “your teacher” rather than “this person.” For example, question 17, which asks, “How often does this person criticize you?” was changed to “How often does your teacher criticize you?” In this research, the companionship and dominance scales were removed as the questions did not pertain to teacher-student relationship and alluded to situations outside the context of the classroom. Twelve questions pertaining to the positive aspect of the teacher-student relationship, closeness, consisted of the first order factors of disclosure, satisfaction, emotional support, and approval, and had reliabilities of .84, .88, .83, and .74, respectively. Twelve questions pertaining to the negative aspect of the teacher-student relationship, discord, consisted of the first order factors of pressure, conflict, criticism, and exclusion, and had reliabilities of .76, .77, .75, and .58, respectively.

Needs satisfaction scale. How well basic psychological needs are met by teachers was measured with the Needs Satisfaction Scale developed by La Guardia et al., (2000) (see Appendix I). The original scale was developed to measure need satisfaction in particular relationships. The needs satisfaction scale in specific relationships is similar to

the scales developed by the authors for needs satisfaction at work and needs satisfaction in general; however, the needs satisfaction scale in specific relationships is a more parsimonious instrument consisting of only 9 items, whereas the others contained 21 items. The scale was initially developed to study how satisfying an individual's need for autonomy, competence, and relatedness affected an individual's attachment to others, and has been modified by others for use in their studies, such as Garn and Wallhead (2015) and Ntoumanis (2005) modifying it for use in physical education.

The needs satisfaction scale, according to the authors, can be used to measure need satisfaction in specific relationships. Questions on the instrument take the form of "When I am with XXXXXXX, I feel free to be who I am" (La Guardia et al., 2000, p. 384). In this research, each question replaced XXXXXXX with "my teacher" to read, for example, "When I am with my teacher, I feel free to be who I am."

The original Needs Satisfaction Instrument contained fifteen items and was based on a 9-point Likert scale, and had reliabilities of .92, .92, .92, and .90 when measuring perceived levels of autonomy, competence, and relatedness in regard to mother, father, romantic partner, and friends in a college level sample. In a second study that was related to their initial study, La Guardia et al., (2000) refined the instrument and removed six items. The final Needs Satisfaction Scale included nine items with three questions each for autonomy, competence, and relatedness rated on a 7-point Likert scale ranging from 1 to 7 (not at all true to very true). Reliabilities were .91, .94, .88, .85, .90, and .90 when measuring perceived levels of autonomy, competence, and relatedness in regard to mother, father, romantic partner, best friend, roommate, and other significant adults. The results of confirmatory factor analysis of the items indicated a three factor model with

adequate fit (RMSEA = .10, CFI = .96). The results of the Chi-square analysis supported a three factor model for measuring needs satisfaction.

The nine items, three for each psychological need, were collected for each individual. Questions four, six, and nine were worded in the negative direction and were reverse scored by subtracting from 8. Satisfaction of each of the basic psychological needs was identified by larger values.

Classroom engagement inventory (CEI). Classroom engagement was measured with the CEI developed by Wang et al., (2014). The CEI was developed to fill the gap in available resources to study engagement and is the sole inventory that measures the multiple dimensions of engagement at the classroom level. The instrument has been validated for use in grades four to twelve and was found by Wang et al., to be invariant across age, class, gender, and socioeconomic status. Other than Wang et al. creating the inventory, the CEI has been used in only one correlational study in Turkey (Sever et al., 2014).

The CEI is a 5-point Likert scale measure based on frequency ranging from 1 to 5 (never to daily) with higher scores indicating more engagement. The original version of the CEI contained 24 questions and measured five dimensions of engagement that include affective engagement, behavioral engagement (compliance), behavioral engagement (effortful class participation), cognitive engagement, and disengagement with McDonald’s omega of .90,.82, .82, .88, and .82, respectively. Sever et al., (2014) reported Cronbach Alpha’s of .87, .82, .74, .89, and .69 for each of the dimensions.

Disaffection/disengagement was not studied in this research; therefore, questions nine, twelve, and twenty-one were removed. To attain a more parsimonious survey

instrument for participants to complete, other questions on the CEI have been removed based on factors loadings reported by Wang et al. (2014). While considered high loadings, questions ten and twenty with factor loadings of .78 and .68, respectively, were removed from affective engagement. Two questions, one and five with factor loadings of .58 and .67, respectively, were removed for behavioral engagement (effortful class participation). Questions thirteen, eighteen, twenty-two, and twenty-four with factor loadings of .58, .68, .59, .64, respectively, were removed from cognitive engagement. Behavioral engagement (compliance) was not altered, as there were only three questions pertaining to this dimension. A total of thirteen items remained to measure affective, behavioral (effortful class participation), behavioral (compliance), and cognitive engagement.

Student outcomes. Norm-referenced status scores, scale status scores, and growth scores on the Georgia Milestones assessment for the 2015-2016 school year, student yearlong GPAs, and student's fourth term averages were obtained from the curriculum department and merged with student responses using the student ID as the primary key. The norm-referenced status, scale, and growth scores were determined by the state of Georgia and were unaltered prior to and after the merger with student responses. The norm-referenced and scale score of each assessment, as determined by the state of Georgia, was used as the status score with growth scores being calculated by the state of Georgia.

Data Analysis

This study utilized structural equation modeling (SEM), which is capable of testing complex models of construct relationships with both observed and unobserved

variables (Byrne, 2010; Hoyle, 2014; In'nami & Koizumi, 2013), which cannot be analyzed by regression models (In'nami & Koizumi, 2013; Teo, Tsai, & Yang, 2013). SEM is a confirmatory process whereby a hypothesized model of relationships is tested against observed data from a sample population (Nachtigall, Kroehne, Funke, & Steyer, 2003) while also measuring direct and indirect effects of exogenous and endogenous variables (Hoyle, 2014; Teo, Tsai, & Yang, 2013).

Statistical analysis of data which included descriptive statistics, reliability coefficients, univariate and multivariate outliers, and collinearity coefficients were completed using Statistical Package for Social Sciences (SPSS) version 23 and confirmatory factor analysis and structural equation modeling were completed using Analysis of Moment Structures (AMOS) version 23, with both analyses utilizing an alpha level of $\alpha = .05$ for all estimations. The dataset was screened for abnormalities, missing data, and inconsistencies. Response sets with zero standard deviation were removed as these students chose the same response throughout the survey and provided zero variability and should be removed according to In'nami & Koizumi, 2013.

Missing data, due to item non-response, may bias data and impact effect sizes (Garson, 2015). Garson further stated the two commonly used methods to address missing data are listwise deletion and imputation. Listwise deletion involves deleting the complete case for the respondent while data imputation attempts to estimate the missing value based on the subjects other responses and other similar subjects responses (Byrne, 2010; Garson, 2015; Hoyle, 2014; Kline, 2011). According to Garson, "If missingness is due to unmeasured variables related to the dependent variable, data are MNAR and should not be imputed" (p.16); therefore, listwise deletion was the preferred method in

missing norm-referenced status, scale, and growth scores and the corresponding datasets were removed.

Univariate and multivariate outliers were then addressed. To address univariate outliers, parcels were created for the indicators of closeness, discord, and engagement as the parceled indicators were ultimately used in the structural equation model. Z-scores were calculated and responses that had z-scores greater than |3.29| were removed (Tabachnick & Fidell, 2013). Multivariate outliers have extreme values on two or more variables (Byrne, 2010) and were identified using Mahalanobis distances (Byrne; In'nami & Koizumi, 2013; Kline, 2011). For each measurement model and the final structural equation model, the Mahalanobis distance chi-square test statistic was determined by using the number of observed variables in each model as the degrees of freedom at a significance level of p = .001. Responses were considered multivariate outliers and were removed if they were both greater than the chi-square critical value and had a Mahalanobis distance value vastly different from other responses (Tabachnick & Fidell, 2013).

SEM/CFA results are heavily influenced on the assumption of univariate and multivariate normality as violation of this assumption leads to inflated chi-square test statistics and underestimated standard errors for parameter estimates (Hoyle, 2014; Newsom, 2005) if maximum-likelihood estimations methods are used. While Kline, Newsom, and In'nami and Koizumi (2013) stated multivariate normality can be assumed if univariate distributions are normal, Byrne (2010) noted that it is still possible to have non-normal multivariate distributions if univariate distributions are normal. All univariate distributions for both individual items and parceled indicators were analyzed

for normality through inspection of histograms and evaluation of skewness and kurtosis test statistics (Kline, 2011).

According to Newsom (2005), skewness values greater than |2| and kurtosis greater than |7| indicate a variable is non-normal with kurtosis values being more important than skewness. In'nami and Koizumi (2013) indicated that values exceeding |3| and |21| for skewness and kurtosis values respectively are extremely non-normal while a skewness of |2| and a kurtosis of |7| is moderately non-normal. Maximum-likelihood estimation methods assume no excessive kurtosis of observed variables (Hoyle, 2014). While not a strict cutoff, |2| and |7| was used to judge whether or not skewness and kurtosis data respectively was satisfactory to implement maximum-likelihood estimation methods. Multivariate normality was assessed using Mardia's normalized estimate (Byrne, 2010). If inspection of histograms, skewness and kurtosis values, or Mardia's normalized estimate indicated possible violations of univariate or multivariate normality, the alternate Bayesian estimation method was implemented to verify maximum-likelihood estimations as maximum-likelihood estimates may be problematic.

Measure validation and measurement model testing. The inventories used in this study were modified from the original versions and descriptive statistics indicated possible issues with the measurement instruments, therefore confirmatory factor analysis (CFA) was conducted on each of the measurement instruments (NRI, BPNS, and CEI) to validate each of the measures prior to moving forward with structural model estimation. Confirmatory factor analysis was completed using individual items and not parcels and was guided by model fit indices, factor weights, modification indices, and standardized residual covariances. In the models, each indicator item was assumed to be a continuous

measure of the indicator itself along with an amount of error, the error terms are independent of each other, and the association between factors is unanalyzed (allowed to covary) (Kline, 2011). A single factor loading per factor was set to one to assign a scaling metric to allow for computation of variance and covariances.

Items retained after measure validation were used to create parcels for the NRI-RQV and CEI to simplify the hypothesized model, create a more continuous measurement scale, and to alleviate the issue of needing an unobtainable number of student responses. It is possible to consolidate two or more indicators of a construct into a single composite or parcel indicator by summing or averaging the items (Bandalos & Finney, 2001; Hoyle, 2014; Little et al., 2002; Cunningham, Shahar, & Widaman, 2002), which was done with the latent constructs of closeness, discord, and engagement. According to Little et al., item-level data tend to have lower reliability, lower commonality, and a greater likelihood of non-normal distributions. When using composites, Hoyle, and Bandalos and Finney stated each composite indicator is more reliable than the individual item. Little et al. indicated aggregated items are a better indicator of a construct. Hoyle, however, noted that within instrument measurement errors, were not accounted for. With ordinal data, individual items have fewer intervals, whereas parcels have a greater number of intervals (Little et al., 2002), effectively making the data more continuous (Bandalos & Finney, 2001). In terms of model estimation, parcels offer a more parsimonious model, have fewer chances of residuals being correlated and lower amounts of sampling errors, and indices of fit are more acceptable (Little et al., 2002). SEM is a large sample statistical technique that requires large samples based on number of observed variables. By using parcels, fewer

participants can be included in the studies (Bandalos & Finney, 2001). The decision was made by the researcher to utilize parcels for the constructs of closeness, discord, and engagement to simplify the hypothesized model, create a more continuous measurement scale for the teacher-student relationship and engagement inventories, and to alleviate the issue of needing an unobtainable number of student responses.

While parceling is acceptable under a SEM framework, parceling should be done based on specific criteria. According to Little et al. (2002), if the focus of the research and hypothesized model is to understand the relationship between observed indicators, parceling should not be used; however, if the focus of the research and model is to understand the relationship between constructs, parceling is acceptable, which is true of this research. Parceling of items should only be done under situations of unidimensionality because items measuring the multidimensionality of a construct are likely to be multidimensional and are difficult to parcel (Little et al.). Bandalos and Finney (2001) stated that individual items must be valid for the construct the items are measuring and must be unidimensional and not load on other factors. Both Hoyle (2014) and Little et al. indicated that items on an inventory can be parceled based on the inventory subscales, so long as the factorial structure is supported by prior research, which is true of the measures used in this study. CFA was completed on the measurement models of NRI-RQV, BPNS, CEI, and outcome following the same procedure for measure validation with factor score results imputed into SPSS.

Model specification. SEM was implemented to analyze and estimate model fit indices, model errors, and model parameters of the hypothesized model through use of Analysis of Moment Structures (AMOS) version 23. The following steps were used to

conduct the SEM analysis: model specification, identification, estimation, testing, and modification. A structural model was hypothesized a priori identifying the relationship between manifest and latent variables based on prior research (Byrne, 2010; In'nami & Koizumi, 2013; Kline, 2011; Teo, Tsai, & Yang, 2013) and included the previously validated measurement instrument indicators. The initial hypothesized model was comparable to that of the Self-Systems Process Model, was linear in nature (see Figure 3), and utilized the results of the measurement models. While there was support for feedback loops between basic psychological needs and engagement, those loops were not investigated in this research. The model was recursive in that it did not include feedback loops, and all causal effects were unidirectional (Byrne; Hoyle, 2014).

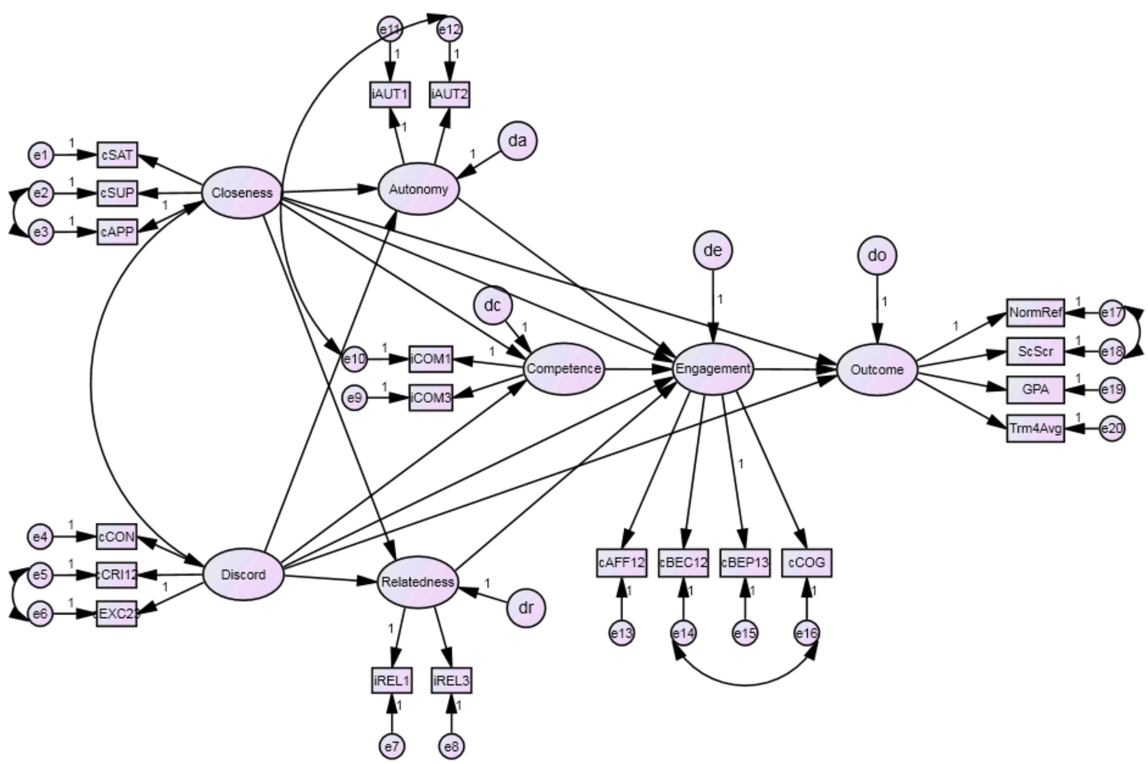


Figure 3. **Hypothesized structural model of the impact of closeness/discord on satisfaction of students' basic psychological needs, student engagement, and student outcomes.**

In a SEM model, squares/rectangles represent observed variables (In'nami & Koizumi, 2013) and are called indicators, which can be continuous or categorical (Hoyle, 2014). Circles/Ovals represent latent variables and are assumed to be continuous (Hoyle). Endogenous variables are dependent variables and exogenous variables are independent variables with endogenous variables being influenced by exogenous variables (Byrne, 2010) and have arrows pointing to them from exogenous variables (Hoyle).

There were seven latent variables in the proposed model, represented by ovals and included closeness, discord, autonomy, competence, relatedness, engagement, and outcome. Each construct was measured by the corresponding indicators represented by squares/rectangles and had corresponding error terms represented as circles (Hoyle, 2014).

As a general rule in SEM, Hoyle (2014) recommended having at least three indicators to identify latent variables as SEM takes into account random or measurement error of latent variables (Petrescu, 2013). Petrescu further stated that, typically, as the number of items used to measure a construct increases, measurement error decreases and reliability increases. Many researchers advise multiple indicators of a construct to improve the validity of measurement and to correct measurement errors (Bergkvist, 2015). All latent variables in the proposed model initially had the minimum number of three indicators to identify latent variables.

All indicators in the final model had an associated error term (circle) that identified each indicator had non-random or measurement error (Byrne, 2010). Byrne also stated that prediction of endogenous terms by exogenous terms is not without error and, therefore, residual error or disturbance terms were included. The five latent

endogenous variables of autonomy (Da), competence (Dc), relatedness (Dr), engagement (De), and outcome (Do) had a corresponding disturbance term. Closeness and discord had an associated variance term and were also set to covary. One indicator per latent variable, along with all error terms, were fixed to 1 to set the scaling metric of the factor (Hoyle, 2014; Tabachnick & Fidell, 2013).

The process of model specification was completed by developing the hypothesized model a priori based on the literature review (Kline, 2011). Closeness and discord, measures of the teacher-student relationship (context) were hypothesized to affect autonomy, competence, and relatedness (self), which all influence engagement (action) which consequently influences student growth scores, status scores, GPA, and term average (outcome).

Model identification. In order for AMOS to estimate a unique value for every parameter in the model, the model must be identified and have a degrees of freedom greater than zero (Kline, 2011). If a model is identified, there are enough observed indicators to estimate unknown parameters in a model (Nachtigall et al., 2003). In other words, according to Hoyle (2014), there must be more known information than unknown. When a unique value for each parameter can be obtained using the covariance matrix, the model is identified (Hoyle), which, in the case of the hypothesized model, is true. The number of freely estimated parameters was determined by adding factor loadings, path coefficients, error variances, disturbances, variances, and covariances. Degrees of freedom is a function of the number of observed variables in the model and the number of elements in the correlation matrix (Hoyle, 2014). The number of elements in the correlation matrix was calculated using the equation $p\ (\ p + 1)\ /\ 2$ where p was the

number of observed variables. Degrees of freedom, the difference between known and unknown information (Hoyle, 2014), was determined by subtracting the number of free parameters from the number of elements in the correlation matrix.

Model estimation.

Sample size. SEM is a large sample technique requiring a large number of responses with a minimum sample size of 200 (Kline, 2011). Teo et al., (2013) and In'nami & Koizumi (2013) recommended that the sample size be equal to 10 participants per parameter estimated. If the observed data is not normal, sample size should be increased to 15 participants per parameter (Teo, Tsai, & Yang, 2013). The more complex the model, the larger the sample size is required (Kline, 2011; Teo, Tsai, & Yang, 2013). Based on the final model, a sample size between 200 and 630 was recommended by the literature with that latter being better in the situation that the model was complex and data non-normal.

Multicollinearity. Constructs and indicators were checked for signs of multicollinearity. Multicollinearity, a result of high correlations between measures of different constructs or indicator variables, can occur when highly related independent variables are used to predict a dependent variable (Bagozzi & Yi, 2012; Byrne, 2010; Garson, 2012; Kline, 2011; Larwin & Harvey, 2012; Tabachnick & Fidell, 2013). Multicollinearity can cause many issues to arise in structural equation modeling such as non-convergence, negative variance estimates (Heywood Cases), biased parameter estimates, parameter estimates with unexpected impact and improper signs (Bagozzi & Yi, 2012), inflated standard errors of the collinear variables, and insignificant findings of variables due to inflated standard errors (Garson, 2012; Larwin & Harvey, 2012). If

multicollinearity is an issue, beta weights and R-squares are not reliable, large standard errors can occur, and there is diminished power; however, estimates are unbiased (Garson, 2012). When multicollinearity occurs between two independent latent variables that are then used as a cause of another dependent variable, standardized regression weights could be greater than |1| along with large standard errors for the unstandardized regression weights for the latent variables that are multicollinear (Garson). Multicollinearity may also result in high covariances between the offending variables. Possible multicollinearity may result when two independent indicators or factors have intercorrelations greater than .80, a variance inflation factor (VIF) greater than 4, and/or a tolerance less than .20 (Garson). The tolerance statistic takes into account the interaction effect of other independent variables as well as correlations between the variables (Garson).

Imputed factor scores for closeness, discord, autonomy, competence, relatedness and engagement and indicators scores were used to estimate bivariate correlations, tolerance scores, and variance inflation factor statistics. According to Kline, (2011), In'nami & Koizumi, (2013), and Tabachnick & Fidell (2013), multicollinearity can be adjusted for by deleting or combining redundant variables. If multicollinearity occurred in the model, indicators and factors were adjusted as needed and the model was respecified.

*Estimation methods***.** The purpose of estimation in SEM is to determine how well the hypothesized model fits observed data (Byrne, 2010). According to Byrne, multiple studies conducted with SEM over the past 15 years revealed most to include Likert-type scales incorporating the maximum-likelihood (ML) estimation procedure. ML model

estimation in SEM is the most widely used estimation method and is a normal theory method assuming multivariate normality, continuous variables, and no missing values in the raw data file (Hoyle, 2014; Kline, 2011; Morata-Ramirez & Holgado-Tello, 2013; Rhemtulla et al., 2012). Lei (2009) and Hoyle further added there should be no excessive kurtosis. ML estimation incorporates Pearson correlations, which assumes continuous measurement (Morata-Ramirez & Holgado-Tello, 2013), typically utilizing an interval measurement scale (Holgado-Tello, Chacón–Moscoso, Barbero–García, Vila–Abad, 2010).

The data set contained within this research contained multiple Likert-scale inventories, two with 5 categories, one with 7 categories, and were ordinal. Lei (2009) and Morata-Ramirez and Holgado-Tello (2013) recommended, when using ordinal variables that exhibit normality such as Likert-scale inventories, polychoric or polyserial correlation estimation methods. Kline (2011) further added that other estimation methods should be used if the data does not satisfy assumptions of normality.

Treating ordinal data as continuous has led to biased parameter estimates, incorrect standard errors, and incorrect model test statistics, especially when there were few categories (Rhemtulla et al., 2012). When treating ordinal data as continuous, use of Pearson correlations in the ML estimation method can lead to faulty estimates (Morata-Ramirez & Holgado-Tello, 2013). Pearson correlations are higher between continuous variables and tend to underestimate the strength of association between ordinal variables (Byrne, 2010; Holgado-Tello et al., 2010; Lei, 2009) and underestimates factor loadings, especially with four or fewer categories (Byrne; Hoyle, 2014; Rhemtulla et al., 2012). Polychoric and polyserial correlations have been found to perform better than Pearson

correlation in multiple studies that included ordinal variables; however, polychoric and polyserial correlations tend to produce biased estimates when normality assumptions are violated or with low sample size (Lei). Polychoric correlations have less bias than Pearson correlation when working with ordinal data and Morata-Ramirez and Holgado-Tello found that polychoric correlations were more advantageous than Pearson correlations when dealing with ordinal variables.

Many researchers that included Likert inventories with five or more categories in their research have treated the variables as continuous with little difference in findings when treating the variables as ordinal (Newsom, 2005). According to Dollan (1994), Hoyle (2014), and Rhemtulla et al. (2012), in order to apply factor analytic theory to Pearson correlations, at least five response categories are required. The premise behind treating ordinal data as continuous is that as more categories are included, the data becomes more continuous (Rhemtulla et al.). The underestimation of association between observed ordinal variables and negative bias disappears when the number of categories reaches five (Rhemtulla et al.).

While Rhemtulla et al. (2012) indicated that treating ordinal data as continuous can lead to biased results when there are few categories, as there are more categories, errors in estimation decrease. Rhemtulla et al. investigated how many categories were necessary in order for the ML estimation method to produce reliable and valid results for ordinal data by comparing the categorical least-squares estimation method with robust corrections to the ML estimation method with robust corrections. The researchers found that categorical least squares was slightly less accurate than ML with five to seven categories, but was found to be more accurate than ML with two to four categories. With

two categories, factor loadings were substantially negatively biased; however, once five categories was reached, the relative bias was less than 10%, which is acceptable. Both methods were more accurate with a greater number of categories and had comparable standard error estimates leading Rhemtulla et al. to conclude, that having at least five categories allowed for the use of continuous robust ML so long as categories were roughly normal. When ML estimation methods are used with ordinal variables, it is advised to apply robust corrections as proposed by Satorra and Bentler (1994).

Byrne (2010) found that the chi-square test statistic was influenced very little by ordinal data so long as underlying data is approximately normal. In situations of differential skewness, implying the data is not normal, the chi-square test statistic becomes inflated, standard errors tend to be lower, and error variance estimates become unreliable. Without stating a number of categories, Byrne stated that so long as the number of categories is large and observed data is normal, ML can be used as the estimation method in SEM analysis.

For ordinal data, multiple recommendations have been made in the literature as to the estimation method to be used such as polychoric or polyserial correlations (Lei, 2009), categorical least squares (Rhemtulla et al., 2012), weighted least squares (In'nami & Koizumi, 2013) and the Bayesian approach, which uses the Markov Chain Monte Carlo (MCMC) approach (Newsom, 2005). The current research was conducted using the AMOS and SPSS version 23 statistical package and is limited in the approaches to estimation methods of model parameters with ordinal data. The case has been made for use of the ML estimation method under satisfaction of normal conditions; however, the Satorra and Bentler (1994) robust correction method is not available in AMOS version 23

and will not be used. Closeness, discord, and engagement were measured using 5 point Likert-scale with items being parceled while autonomy, competence, and relatedness were measured using a 7 point Likert-scale inventory, both situations, which support that the observed data can be treated as continuous.

If the ordinal variables were found to violate the assumption of normality, then the Bayesian estimation approach was utilized to test the hypothesized model. Bayesian estimation has been used with and produced reliable results for ordinal data (Muthén & Asparouhov, 2011). Bayesian estimation is a computation tool using an iterative process known as the Markov chain Monte Carlo (MCMC) (Newsom, 2005) to obtain parameter estimates by sampling the observed set of data (Hoyle, 2014). Whereas the focus of ML estimation is on fitting the covariance matrix of the hypothesized model to the sample covariance matrix of the observed data (Hoyle), Bayesian estimation focuses on the use of the raw observations (Song & Lee, 2012). ML estimation requires underlying assumptions of normality and large sample sizes (Byrne, 2010; Hoyle; Kline, 2011; Muthén & Asparouhov), whereas Bayesian estimation can produce reliable estimates with non-normal data, small sample sizes, and skewed distributions (Muthén & Asparouhov; Song & Lee).

Model testing**.** Following the completion of model estimation, model fit was evaluated (Byrne, 2010; Hoyle, 2014; Kline, 2011), and parameter estimates were interpreted to determine how well the model fit the observed data. Assessing model fit entailed evaluating goodness of fit indices and parameter estimates. Parameter estimate values should be feasible, standard errors should be appropriate, and statistical significance of parameter estimates should be evaluated (Byrne, 2010). Byrne further

stated that parameter estimates should be in line with the literature and theory and have proper signs such as positive and negative, and they should be in relative agreement in size. Small standard errors indicate accurate estimation, whereas large standard errors indicate problems in that parameters cannot be determined (Byrne).

How well a hypothesized model fits the observed data is measured through use of the chi-squared test statistic and goodness of fit indices (Byrne, 2010; Hoyle, 2014; Kline, 2011; Lei, 2009). Acceptable fit indices indicate that the model is supported by the data (Nachtigall et al., 2003). The chi-square test statistics is an absolute fit index, which when non-significant, indicates the hypothesized model fits the data (In'nami & Koizumi, 2013; Teo, Tsai, & Yang, 2013); however, chi-square tends to reject the true model more often than not and tends to favor small samples (Hoyle, 2014). Chi-square is not a good indicator of fit unless comparing to competing models and has been found to be sensitive to sample size and cannot be used as a sole indicator of model fit (Teo, Tsai, & Yang, 2013); therefore, the goodness of fit of the model was assessed through evaluation of multiple indices (Kline, 2011). Kline elaborated that fit indices are a measure of the overall fit of the model, but are susceptible to poorly fitted parts of the model, and multiple goodness of fit indices should be identified when evaluating the hypothesized model. Other than the chi-square test statistic, fit indices lie in the range of 0 to 1.0.

The goodness of fit of the proposed model was evaluated utilizing the Steiger-Lind Root Mean Square Error of Approximation (RMSEA), Bentler Comparative Fit Index (CFI), Joreskog-Sorbom Goodness of Fit Index (GFI), and the Tucker-Lewis Index (TLI) (Kline, 2011). RMSEA is sensitive to model size and performs more favorably

with larger models and data that is not normal. RMSEA tends to have lower values with larger sample sizes and degrees of freedom. Lower values indicate better model fit with values lower than .05 (Teo, Tsai, & Yang, 2013), indicating good fit and values from .05 to .10, indicating acceptable fit and values greater than .10 indicating poor fit (Kline). The CFI is an incremental fit index that compares the hypothesized model to the independence model. Unlike RMSEA, larger values indicate better fit with with good fit values being .95 or greater (Kline). The GFI is an absolute fit index that compares the hypothesized model to no model at all by assessing the relative amount of observed variance and covariance is explained by the model (Teo, Tsai, & Yang). Similar to CFI, larger values indicate better fit with greater than .95, representing good fit (Kline; Teo, Tsai, & Yang), and model complexity does not influence the test statistics (Kline). The TLI is an incremental/absolute fit index that compares the x model to the y model. Values greater than .95 indicate good fit (Kline, 2011).

According to the literature, the indicators previously listed work well with both continuous and categorical data. Newsom, (2005) found that RMSEA, TLI, and CFI performed reasonably well with categorical model estimation while Holgado-Tello et al. (2010) found that GFI, AGFI, and RMSEA, were generally in agreement whether using polychoric or Pearson correlation estimation methods, with polychoric being slightly more accurate.

Model respecification. The model was respecified to better fit the data, but was done within limits to not over fit the model to the data, as extensive model respecification may result in a model that is not generalizable to other schools (Kline, 2011; Teo, Tsai, & Yang, 2013). Model respecification took into account and was justified based on prior

research, along with the data provided by AMOS. AMOS provided two sets of statistics that assist in model respecification, which included modification indices and standardized residual covariances. The modification indices table provided a parameter estimate of improved chi-square value of fixed parameters of covariances and regression weights if they are reestimated freely. The modification indices of regression weights were not addressed; however, modification indices for error variances were addressed. Theoretically justifiable and substantive modification indices of covariances were used to improve model fit by setting significant covariances to freely estimate. According to Byrne (2010), standardized residual covariances greater than |2.58| are statistically significant and identify areas of model of misspecification and were addressed. The final step of model respecification was to remove non-significant paths because Byrne indicated that non-significant parameters should be deleted from the model.

Model estimation with growth. Scale Score (ScScr) was removed from the final structural model and substituted with student growth. Prior to estimation, the data were screened for multivariate outliers. The results of the analysis were compared against the ScScr model.

Multigroup Invariance

In SEM, it is possible to analyze multiple groups simultaneously to determine if the model is equivalent across groups (Byrne, 2004; Hox & Becher, 2004). Multigroup testing was utilized to address research question two: To what extent is the effect of teacher-student relationships on student growth percentiles invariant across low socioeconomic status students and non low socioeconomic status students, and to what extent is the effect of teacher-student relationships on student growth percentiles

invariant across white students versus non-white students? Evidence of invariance across groups was based on the chi-square difference test between the configural model and the model being tested (Byrne, 2010). The difference in chi-square and degrees of freedom was compared to Table C.4 in Tabachnick and Fidell (2013) at $\alpha = .05$. If the calculated chi-square difference was greater than the critical value listed in table C.4, it was statistically significant and indicated groups were equal.

The modified hypothesized model was set as the baseline model for multigroup testing, and all groups were tested simultaneously against the same model and was termed the configural model using AMOS (Byrne, 2004). As a first step in testing equivalency, Byrne (2004) recommended to test the fully constrained model by constraining all factor loadings, factor variances, and factor covariances equal across the groups. This research was not concerned with constraining error variances or error covariances because Byrne and Byrne (2010) indicated this is too restrictive of a test for equality. If the full model was found to be invariant across the groups, no further testing was completed; however, if not equivalent, Byrne stated that invariance testing commence with measurement models first, followed by the structural model.

To assess metric invariance of the measurement models, factor loadings for all indicators of the measurement models were constrained equal and estimated to determine if groups were invariant (Templin, 2012). If results indicated group differences, each of the measurement models was individually tested to determine which indicators were not invariant across groups. To do this, only the construct under study had the indicators constrained equal, while the other measurement models were freely estimated. Individual indicators of the non-equivalent constructs were then tested by constraining each

indicator and freely estimating the others, constraining invariant indicators and freely estimating non-invariant indicators in each subsequent test until all indicators of construct had been assessed. Structural invariance of the full model was then tested by constraining factor variances and covariances, leaving all factor loadings constrained that were found to be invariant (Byrne, 2010; Templin, 2012) in a process similar to measurement model testing.

CHAPTER IV
RESULTS

Introduction

The purpose of this quantitative research was to examine the extent that teacher-student relationships influenced basic psychological needs, engagement, and growth/status scores using the SSPM as a framework, with outcome being measured using the Georgia Milestones standardized assessment norm-referenced scores, scale scores, yearlong class GPA, fourth term student averages, and student growth percentiles.

The research was guided by the following three research questions:

1. To what extent does the teacher-student relationship influence satisfaction of basic psychological needs which influence engagement and, consequently, influence student growth percentiles as compared to student status scores using an identical methodological setup (Context → Self → Action → Outcome)?
2. To what extent is the effect of teacher-student relationships on student growth percentiles invariant across population subgroups? (i.e. Low socioeconomic status students versus high socioeconomic status students and White students versus non-white students)
3. To what extent does the teacher-student relationship influence level of student engagement (Context → Self → Action)?

Data Screening

Responses were entered into SPSS (*n* = 543) and screened for missing values. Eight students lost internet connection when completing the survey on iPads, and one or more whole sections of question responses were missing. Since responses to entire sections of the NRI, BPNS, and/or CEI were missing, imputation would have been impossible, and the eight corresponding response sets were deleted (*n* = 535). Based on how the data collection instrument was set up and administered to students, all other successfully completed surveys had no missing responses; however, there was one student who did not participate in the 2015-2016 Georgia Milestones assessment, along with eleven students who did not participate in the 2014-2015 Georgia Milestones assessment due to being from another country, state, or private school; therefore, 2015-2016 growth score could not be calculated.

According to Garson (2015), "If missingness is due to unmeasured variables related to the dependent variable, data are MNAR and should not be imputed" (p.16); therefore, listwise deletion was the preferred method if a response set was missing norm-referenced status, scale, and/or growth scores. Twelve students and their 23 (4.3%) corresponding response sets were removed from SPSS (*n* = 512). The data were then examined for inconsistencies to identify students that had no deviation in responses, and only one score met these criteria (case 238); however, the student response was already removed from the dataset due to not having growth score.

Participants in the research were representative of the school and district population because race and socioeconomic status percentages were nearly identical. More seventh grade students participated due to more seventh grade teachers volunteering to participate in the research and owning the parental consent forms. ELA, math, science, and social studies had near equal distribution of responses; however, there were more responses by females than males (See Table 3).

Table 3

Participant demographics

Demographic	Category	Count	Percentage
Grade			
	7	315	61.5
	8	197	38.5
Gender			
	Female	312	60.9
	Male	200	39.1
Race			
	White	373	72.9
	Black	90	17.6
	Hispanic	23	4.5
	Multiracial	24	4.7
	Asian	2	.4
Socioeconomic Status			
	Low SES	168	32.9
	High SES	344	67.1
Subject			
	ELA	126	24.6

Math	131	25.6
Science	117	22.9
Social Studies	138	27.0

n = 512

To address univariate outliers, parcels were created for the indicators of closeness, discord, and engagement because these indicators were ultimately used in the structural model. Z-scores were calculated for the parceled items of disclosure (cDIS), satisfaction (cSAT), support (cSUP), approval (cAPP), pressure (cPRE), conflict (cCON), criticism (cCRI), exclusion (cEXC), affective engagement (cAFF), behavioral engagement compliance (cBEC), behavioral engagement participation (cBEP), cognitive engagement (cCOG), the nine individual indicators of basic psychological needs (BPNS), and the five individual indicators of outcome. Responses to survey items were considered univariate outliers if z-score values were greater than |3.29| (Tabachnick & Fidell, 2013). There were 30 response sets that had univariate outliers on one or more factors and were removed from the dataset (see Table 4).

Table 4

Univariate outliers

	cDIS	cCON	cCRI	cEXC	cBEC	GPA	Scale Score
Number of Outliers	12	4	8	2	11	1*	3*
Mean	1.516	1.666	1.624	1.784	4.440	86.60	531.63

Notes	All positive outliers

* Outliers not removed.
Conflict, criticism, and BEC had responses sets with overlapping outliers
n = 482 after removal of univariate outliers.

Preliminary means identified that students tend not to disclose information to teachers; however, twelve students responded that they do. Three parceled factors of discord, cCON; cCRI; and cEXC all had low means, indicating a positive teacher-student relationship (TSR); however, eleven students indicated on one or more of the factors high levels of discord. Eleven students responded lower than their peers on cBEC. The indicators of outcome had four outliers, one for GPA (case 60) and three for scale score (cases 346, 700, 724), which were not removed from the data set because they were within the expected range of acceptable scores. The other parcels of cSUP, cSAT, cAPP, cPRE and the individual indicators of BPNS had no outliers. Univariate outliers were removed (30 cases, 5.9%) from the data set (n = 482).

Multivariate outliers have extreme values on two or more variables (Byrne, 2010) and were detected through the use of the Mahalanobis distance chi-square test statistic (Byrne; In'nami & Koizumi, 2013; Kline, 2011). Measurement instruments were validated using confirmatory factor analysis (CFA) with individual indicator items. Prior to conducting CFAs, multivariate outliers were removed from the dataset for each model using a degrees of freedom equal to the number of observed variables at p = .001 to determine the chi-square critical value test statistic (see Table 5). Responses were considered multivariate outliers and were removed if they were both greater than the chi-

square critical value and had a Mahalanobis distance value spread apart from other response sets (Tabachnick & Fidell, 2013).

Table 5

Multivariate outliers for individual measurement models and full structural model

Measurement Instrument	df	χ2 critical value*	Range of Mahalanobis Distances	Number of deleted responses	Responses remaining
Closeness & Discord	22	48.268	120.5 to 67.0	10	n = 472

Table 5 Continued

Measurement Instrument	df	χ2 critical value*	Range of Mahalanobis Distances	Number of deleted responses	Responses remaining
Basic Psychological Needs	9	27.877	39.4 to 32.0	5	n = 467
Engagement	13	34.528	72.6 to 72.5	2	n = 465
Outcome	4	18.467	25.4 to 25.0	2	n = 463
Full SEM Model	20	45.3	63.6 to 45.2	8	n = 455
Outcome with Growth	4	27.877	48.314 to 41.938	4	N = 451

χ2 test statistic determined at $p = .001$

Descriptive Statistics and Normality Assessment - NRI

Student responses (see Table 6) indicated a positive teacher-student relationship (TSR) with higher means on the closeness scale (between "sometimes" and "often feel

this way in class") and lower means on the discord scale (between "never" and "seldom feel this way in class"). The parcel cDIS (*M* = 1.45) had a mean well below the other measures of closeness (M_{cSAT} = 3.64, M_{cSUP} = 2.02, M_{cAPP} = 3.21) indicating that students do not disclose close private information or problems to their teachers. Indicators iDIS2 (*sk* = 2.132, *k* = 4.572) and iDIS3 (*sk* = 2.472, *k* = 5.860) for disclosure were both highly positively skewed and leptokurtic while indicator iDIS1 (*sk* = 1.67, *k* = -.208) had only a slight positive skew and was platykurtic. The question for iDIS1, while similar to the other two, had a weaker connotation. The other two questions had stronger wording, stating that students tell their teacher everything and share secrets and private feelings, which could have influenced how students answered. Based on the histograms, skewness and kurtosis values, a majority of the students did not confide their private feelings and secrets to their teachers. This scale, while appropriate as a measure of certain types of relationships, may not have been appropriate to describe the quality of the TSR since it is apparent that seventh and eighth grade students at this school did not report disclosing secrets, private feelings, everything they are going through, and things they do not want others to know to their teachers, which iDIS1, iDIS2, iDIS3 asked.

Table 6

Network of relationships inventory descriptive statistics

Factor	Mean	S.D.	Skewness	Kurtosis
Closeness				
iDIS1	1.67	0.856	0.964	-0.208
iDIS2	1.44	0.816	2.132	4.572
iDIS3	1.25	0.581	2.472	5.860

cDIS	**1.45**	**0.593**	**1.504**	**1.830**
iSAT1	3.65	1.234	-0.596	-0.636
iSAT2	3.67	1.250	-0.626	-0.652
iSAT3	3.59	1.294	-0.567	-0.826
cSAT	**3.64**	**1.163**	**-0.544**	**-0.693**
iSUP1	1.69	1.006	1.382	1.040
iSUP2	2.39	1.299	0.516	-0.863
iSUP3	1.96	1.192	1.086	0.229
cSUP	**2.02**	**0.938**	**0.846**	**0.228**
iAPP1	2.87	1.394	0.046	-1.261
iAPP2	3.39	1.242	-0.350	-0.836

Table 6 Continued

Factor	Mean	S.D.	Skewness	Kurtosis
iAPP3	3.37	1.167	-0.389	-0.652
cAPP	**3.21**	**1.083**	**-0.139**	**-0.893**
Discord				
iPRE1	2.69	1.350	0.215	-1.176
iPRE2	2.26	1.240	0.665	-0.564
iPRE3	1.76	1.125	1.420	1.079
cPRE	**2.24**	**0.985**	**0.483**	**-0.493**
iCON1	1.82	1.069	1.295	0.933
iCON2	1.34	0.818	2.722	7.188
iCON3	1.53	0.945	1.818	2.496
cCON	**1.56**	**0.801**	**1.803**	**2.890**
iCRI1	1.58	0.992	1.806	2.638

iCRI2	1.64	0.925	1.469	1.677
iCRI3	1.32	0.695	2.270	4.569
cCRI	**1.52**	**0.704**	**1.522**	**1.829**
iEXC1	1.36	0.815	2.552	6.370
iEXC2	1.86	1.112	1.227	0.620
iEXC3	1.87	1.185	1.304	0.735
cEXC	**1.70**	**0.815**	**1.216**	**0.748**

n = 455, i denotes individual items, c denotes composites

Means for the pressure factor were elevated above other factors of discord, indicating that students felt pressured by their teachers. Pressure may be a natural part of TSRs, as teachers constantly urge their students to do things they do not want to do, do not like to do, or do things the teacher wants which the inventory items asked about. Like disclosure, while pressure may be appropriate as a measure of certain types of relationships, pressure may not have been appropriate to describe the quality of TSRs.

Visual inspection of histograms of the individual items was mostly not normal, which was expected with this 5- point Likert scale inventory with most items exhibiting a step-like contour (see Figure 4). Approval indicators had a more normal distribution (see Figure 4).

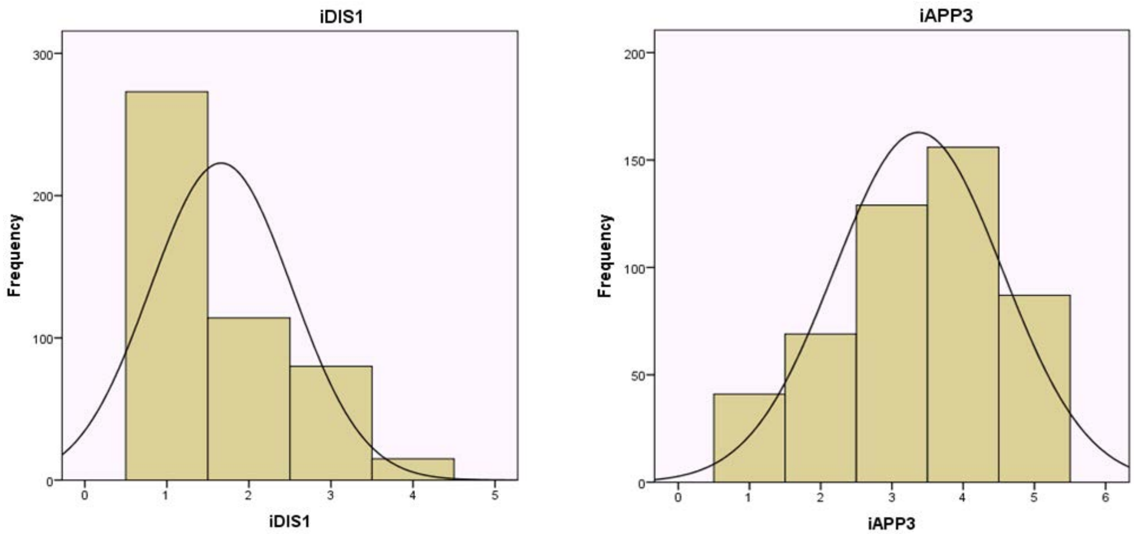


Figure 4. **Sample histograms for closeness and discord**

Histograms of the parceled indicators had a greater number of buckets and had a more normal appearance than individual indicators. The most normal was cAPP followed by cSAT, which had a slight negative skew, and cSUP and cPRE, which were positively skewed. While cSUP was positively skewed, it was not skewed nearly that of cDIS, with responses being more distributed. Being vastly different from the other indicators, cDIS had a majority of parceled responses between 1 and 1.75, indicating most students did not disclose information to their teachers. Histograms of the parceled factors of discord were positively skewed and had the same relative pattern: students strongly felt their teachers did not pressure them, experience conflict, were not criticized by their teachers, or were not excluded from classroom activities.

Descriptive Statistics and Normality Assessment - BPNS

There were nine items, three for each basic psychological need (BPNS). Questions four, six, and nine were worded in the negative direction and were scored by subtracting from 8. Satisfaction of autonomy (M = 4.87), competence (M = 5.20), and

relatedness (M = 4.54) was identified by larger values, and student responses indicated a general satisfaction of their BPNS with the means falling in the range of being somewhat true or higher (see Table 7). Means and skewness and kurtosis values of the reverse worded questions were atypical when compared to non-reverse worded questions. The negatively worded questions of iAUT3R (M = 5.84, sk = -1.357, k = 1.048), iCOM2R (M = 5.92, sk = -1.433, k = 1.490), and iREL2R (M = 5.29, sk = -.934, k = -.107) had significantly higher means, were more negatively skewed, and had higher kurtosis values than the other questions in the corresponding data set. The negatively worded question (iREL2R) did not appear to be as extreme as the other two.

Table 7

Need satisfaction scale descriptive statistics

Factor	Item	Mean	S.D.	Skewness	Kurtosis
	iAUT1	4.61	2.020	-.352	-1.039
Autonomy	iAUT2	4.16	2.055	-.070	-1.214
	iAUT3R	5.84	1.583	-1.357	1.048
	iCOM1	4.80	1.970	-.497	-.891
Competence	iCOM2R	5.92	1.478	-1.433	1.490

Table 7 Continued

Factor	Item	Mean	S.D.	Skewness	Kurtosis
Competence	iCOM3	4.89	1.872	-.541	-.795
	iREL1	4.78	1.995	-.512	-.948
Relatedness	iREL2R	5.29	1.828	-.934	-.107
	iREL3	3.54	2.046	.257	-1.168

n = 455, R indicates the item was negatively worded

Visual inspections of histograms of the items of basic psychological needs were modestly non-normal. Responses were spread over the range from one to seven, with all distributions having the greatest number of responses concentrated about seven. The reverse worded questions, based on responses of non-reverse worded questions, should have had the greatest frequency of responses around one, but also were around seven. The evidence indicated that students struggled to answer the reverse worded questions in an appropriate manner, and these items were further evaluated for inclusion in the validation of the measure through confirmatory factor analysis.

Descriptive Statistics and Normality Assessment - Engagement

In all dimensions of the engagement construct, student responses indicated they were engaged from monthly to daily (see Table 8). All items and composites were negatively skewed. Means of behavioral engagement compliance (BEC) were the highest, while behavioral engagement participation (BEP) means were the lowest. Results for affective engagement (AFF) were consistent across items with similar means, were similarly negatively skewed, and were platykurtic. Item iBEP2 (M = 2.98, sk = -.057, k = -1.449) had a lower mean, was less skewed and was more platykurtic than iBEP1 (M = 3.59, sk = -.570, k = -1.061) and iBEP3 (M = 4.02, sk = -.938, k = -.119).

Table 8

Classroom engagement inventory descriptive statistics

Factor	Item	Mean	S.D.	Skewness	Kurtosis
	iAFF1	3.44	1.399	-0.409	-1.187

Affective Engagement	iAFF2	3.64	1.324	-0.646	-0.827
	iAFF3	3.75	1.296	-0.724	-0.635
	cAFF	3.61	1.206	-0.554	-0.830
Behavioral Engagement Participation	iBEP1	3.59	1.423	-0.570	-1.061
	iBEP2	2.98	1.495	-0.057	-1.449
	iBEP3	4.02	1.135	-0.938	-0.119
	cBEP	3.53	1.060	-0.376	-0.829
Behavioral Engagement Compliance	iBEC1	4.43	0.950	-1.812	2.697
	iBEC2	4.50	0.847	-1.773	2.532
	iBEC3	4.69	0.661	-2.464	6.601
	cBEC	4.54	0.617	-1.487	1.652
Cognitive Engagement	iCOG1	4.04	1.153	-1.100	0.257
	iCOG2	4.42	0.974	-1.789	2.516
	iCOG3	3.80	1.298	-0.772	-0.622
	iCOG4	4.39	0.953	-1.547	1.580
	cCOG	4.16	0.870	-1.052	0.357

n = 455

Behavioral engagement compliance (BEC) questioned how often students listened very carefully (iBEC1), were attentive to things they were supposed to remember (iBEC2), and completed assignments (iBEC3). The item iBEC3 had the largest mean and was the most skewed and leptokurtic distribution (M = 4.69, sk = -2.464, k = 6.601), which influenced the composite results. cBEC consequently had the most skewed and leptokurtic distribution (sk = -1.487, k = 1.652). The composite of cBEC had a total of

eleven univariate outliers, which was heavily influenced by item iBEC3. While iBEC3 was a measure of compliance, the wording and resulting outcome were a different measure of compliance when compared to the other two items in the construct, which was supported by the mean, skewness, and kurtosis results. The item was further evaluated for inclusion in the research through validation of the measure through confirmatory factor analysis. Cognitive engagement items had consistent responses with the exception of iCOG3 (M = 3.80, sk = -.772, k = -.622), which had greater distribution of responses, a significantly lower mean, skewness, and kurtosis. The item iCOG1 stated, “I go back over things I do not understand,” iCOG2 stated, “I think deeply when I take quizzes,” and iCOG4 stated, “If I make a mistake, I try to figure out where I went wrong.” iCOG3 stated, “I ask myself some questions as I go along to make sure the work makes sense,” which is student self-talk and may have been a skill the majority of these students did not have or understand.

No histogram had the appearance of a normal distribution, and all but iBEP2 had peaked distributions to the right, indicating a majority of students were engaged in the classroom. iBEP2 stated, “In this class, I do not want to stop working at the end of class” and was peaked on the left side of the histogram. iBEP1 and iBEP3 stated, “In this class, I form new questions in my mind as I join in class activities,” and “In this class, I get really involved in class activities” respectively, which differed in connotation from iBEP2. Student responses to iBEP2 being peaked to the left were reasonable from an educator's perspective based on the wording of the question. Behavioral engagement compliance and cognitive engagement histograms were stepped similarly to the right while affective engagement and behavioral engagement participation were more

randomly distributed. The composites, while not appearing normally distributed, had a wider range of buckets and took on a more normal appearance.

Descriptive Statistics and Normality Assessment - Outcome

Students' fourth term average (*M* = 87.75, *sk* = -.761, *k* = .229) and yearlong grade point average (*M* = 87.16, *sk* = -.761, *k* = -.069) had similar means, along with a slight negative skewness. Term four average was leptokurtic, GPA was slightly platykurtic, and histograms were close to resembling normal. Scale score (*sk* = .165 and *k* = .639) had fewer students scoring at higher levels and a large number of students scoring in the middle. The histogram appeared to be normal, but was peaked. The means of norm-referenced score (*M* = 69.91) and growth score (*M* = 56.02) were much lower than teacher assigned scores (see Table 9).

Table 9

Student outcomes descriptive statistics

Factor	Mean	S.D.	Skewness	Kurtosis
Term4Avg	87.75	7.999	-.761	.229
GPA	87.16	7.724	-.761	-.069
Scale Score	534.60	46.411	.165	.639
NormRef Score	69.91	24.963	-.875	.027
Growth Score	56.02	28.154	-.255	-1.152

Table 9 Continued

n = 455

Validation of Measures and Measurement Models

Each of the measurement instruments (NRI, BPNS, CEI) used in this research were modified from the original versions, and indicators of outcome were chosen to represent the outcome construct. Modification of inventories, along with descriptive and reliability statistics, indicated there might be issues with the measurement instruments used to develop measurement models. Prior to estimating the structural model, the individual measurement models (inventories) should be validated and psychometrically sound (Byrne, 2010; In'nami & Koizumi, 2013) to ensure accurate findings of the structural model (Byrne) using confirmatory factor analysis (CFA). Confirmatory factor analysis was conducted for each measurement instrument to validate the measure and each measurement model in order to ensure the best possible measurement model in the structural equation modeling analysis.

Original models were specified with subsequent changes made to the model by removing indicators or including covariances, resulting in nested models when compared to the original. According to Brown and Moore (2015); and Byrne (2004) model modification can be verified to have improved fit by utilizing the chi-square difference test to judge whether nested models were significantly better than the original models. The chi-square difference test was used in this research to confirm better fitting models during measurement and structural model modification.

Validation of measure - Confirmatory factor analysis – NRI. A second-order model of NRI, aligned to the findings of Furman and Buhrmester (2009), was created. Closeness and discord were set as second-order factors, while disclosure, approval, satisfaction, support, pressure, conflict, criticism, and exclusion were set as the first order factors, with all individual questionnaire items set to load on the corresponding first-order

factors. In the initial model, the second-order factors of support and exclusion factor loadings were assigned a scaling metric of one, with all scaling metrics being set to one for the third question of each indicator. The initial CFA model was estimated, and as changes were made one at a time, the estimation process was rerun. After initial calculation of estimates, the indicators and latent factors that had the highest unstandardized factor loadings had the scaling metric set to 1 (Byrne, 2010), with estimates calculated again to get the base model statistics (see Figure 5).

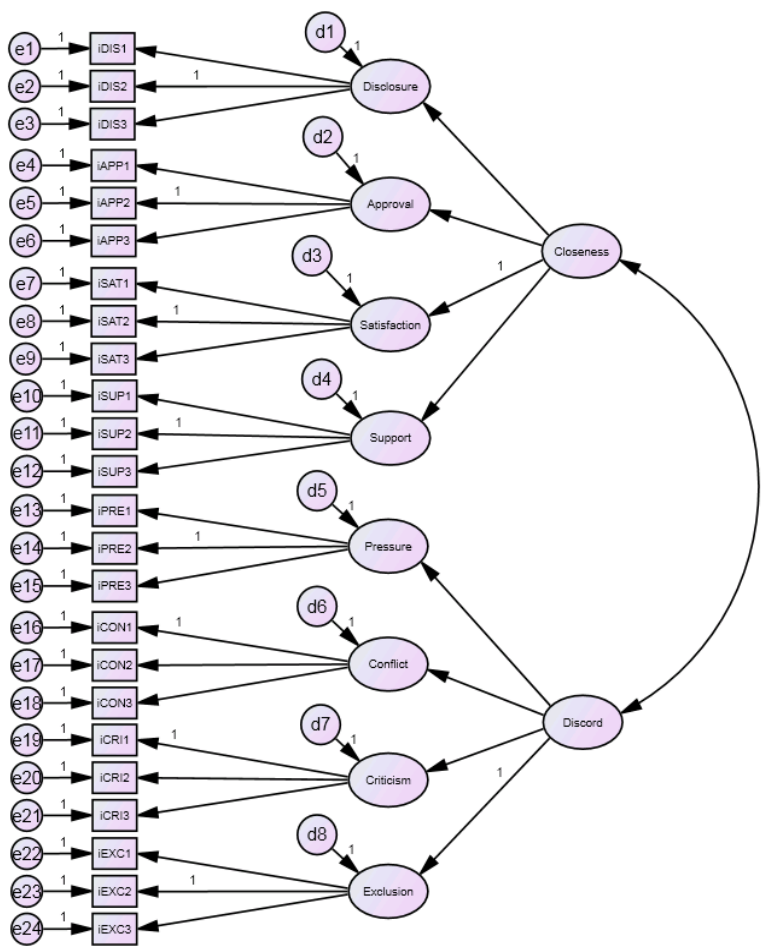


Figure 5. **Second order factor model of closeness and discord**

The initial model showed poor model fit (Model 1, $\chi^2 = 737.1$, $p = .000$, df = 243, GFI = .875, CFI = .900, TLI = .886, RMSEA = .067) (see Table 10). As already detailed, based on student responses, disclosure did not appear to be an acceptable measure of the teacher-student relationship, contrary to the other measures of closeness. Disclosure had means well below the other measures of closeness indicating that students did not “tell your teacher things that you don't want others to know” (q1), “tell your teacher everything that you are going through” (q2), or “share secrets and private feelings with your teacher” (q3). While the disclosure scale may be appropriate in comparing relationships between others as Furman and Burhmester (2009) had intended, this scale may not have been appropriate to describe the quality of the teacher-student relationship since it was apparent that students did not report disclosing personal information to their teachers in this situation. In the initial run of the CFA model, the standardized regression weight of closeness on disclosure ($\beta = .50$) was much lower than approval ($\beta = .94$), satisfaction ($\beta = .90$), and support ($\beta = .75$) with only 24.6% of the variance in disclosure explained by closeness (see Figure 6).

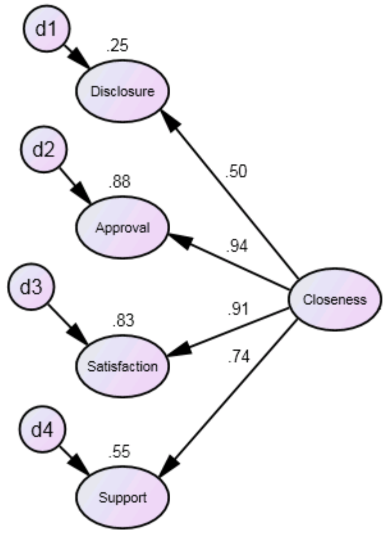


Figure 6. **Standardized regression weights of closeness**

With lower means, dramatic differences in skewness and kurtosis values, and a low standardized regression weight compared to support, satisfaction, and approval, the latent variable of disclosure along with the corresponding indicators were dropped from the model and were not analyzed in subsequent models, which resulted in improved model fit (Model 2, $\chi^2 = 469.2$, $p = .000$, df = 181, GFI = .905, CFI = .935, TLI = .924, RMSEA = .059).

Similar to disclosure, pressure was intended to measure discord and had means different from the other indicators of discord. Pressure means were elevated above other discord factors, which may be a natural part of the teacher-student relationship, as teachers are constantly pushing to get students to be productive in class. In hindsight, this is part of a normal school day experienced by students in that teachers push their students to stay on task, do their work, and excel in what they are doing, which is not a true measure of discord in the classroom. The indicators of pressure and the parceled means were above the other measures of discord, indicating students felt that "your teacher pushes you to do things that you don't want to do" (q1), "your teacher tries to get

you to do things that you don't like" (q2), and "your teacher pressures you to do the things that he or she wants" (q3). In Model 2, the standardized regression weight of discord on pressure (β = .36) was much lower than conflict (β = .72), criticism (β = .86) and exclusion (β = .90). Only 13.0% of the variance in pressure was explained by discord (see Figure 7). While the factors of pressure were not highly skewed or kurtotic, pressure was dropped from the model due to the differing means and low standard regression weight with improved model fit (Model 3, χ^2 = 358.2, *p* = .000, df = 128, GFI = .914, CFI = .943, TLI = ..931, RMSEA = .063).

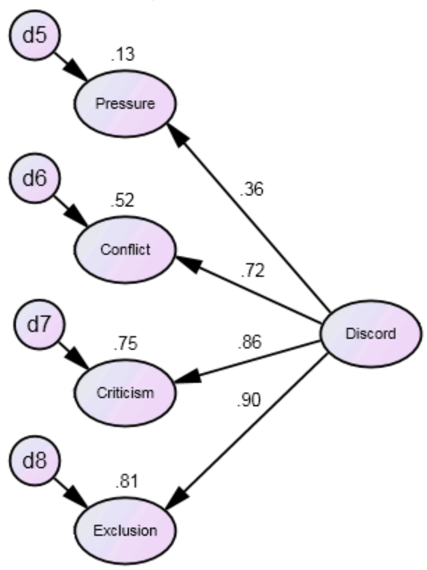


Figure 7. **Standardized regression weights of discord**

Table 10

NRI Measure validation - Model fit indices utilizing individual items

Model	χ^2	df	GFI	CFI	TLI	RMSEA	Notes
1	737.09	243	.875	.900	.886	.067	Original Model with all items
2	469.15	181	.905	.935	.924	.059	Disclosure removed

3	358.22	128	.914	.943	.931	.063	Pressure removed
4	289.39	112	.928	.954	.944	.059	iEXC1 removed
5	243.50	97	.935	.960	.951	.058	iCRI3 removed

The item iEXC1 was more positively skewed (*sk* = 2.54) than iEXC2 (*sk* = 1.22) and iEXC3 (*sk* = 1.30) and was more leptokurtic (*k* = 6.29) than iEXC2 (*k* = .60) and iEXC3 (*k* = .71). The factor loading of exclusion on iEXC1, while statistically significant (β = .44), was lower than iEXC2 (β = .84) and iEXC3 (β = .67) (see Figure 8). Only 19.8% of the variance in iEXC1 was explained by exclusion, and the factor was dropped from the model, which resulted in model improvement (Model 4, χ^2 = 289.4, *p* = .000, df = 112, GFI = .928, CFI = .954, TLI = .944, RMSEA = .059).

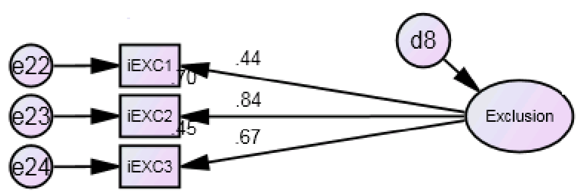


Figure 8. **Factor loadings on first or factor of exclusion**

In the standardized residual covariance matrix, iCRI3 (3.162) had the sole value greater than 2.58, which indicated the item did not fit very well in the model. iCRI3 (*M* = 1.32, *sk* = 2.270, *k* = 4.569) also had a lower mean, greater skewness, and was more leptokurtic than the other indicators of criticism. While the factor loading for iCRI3 (β = .65) was significant and greater than normally accepted values, the indicator was dropped from the model.

The final model adequately fit the data (see Figure 9), with GFI and RMSEA values indicating adequate fit and CFI and TLI values indicating good fit (Model 5, χ^2 = 243.5, *p* = .000, df = 97, GFI = .935, CFI = .960, TLI = .951, RMSEA = .058). The latent

variables of approval (β = .92), satisfaction (β = .93), support (β = .70), conflict (β = .67), criticism (β = .74), and exclusion (β = .89) were statistically significant and had standardized regression weights greater than β = .6. The remaining indicator items of each of the first-order latent variables were statistically significant, had factor loadings greater than .6 and were used to build composites indicators of closeness and discord, which included cSAT, cSUP, cAPP, cCON, cCRI12, and cEXC23 with reliability coefficients of .91, .72, .81, .79, .69, and .73 respectively.

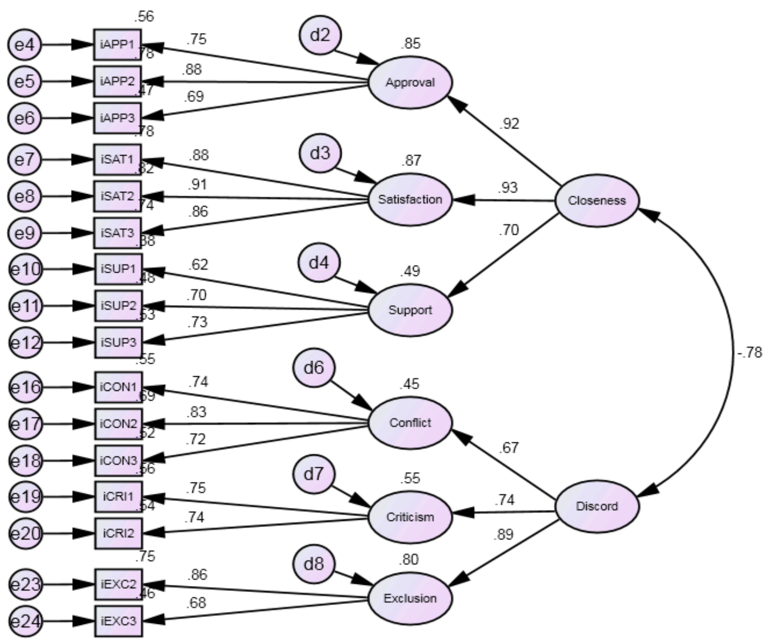


Figure 9. **Measurement instrument validation of closeness and discord**

Validation of measurement model - Confirmatory factor analysis – NRI. The first-order measurement model of closeness and discord consisted of the parceled indicators cSAT, cSUP, cAPP, cCON, cCRI12, and cEXC23. The model had 13 free parameters to be estimated, 21 sample moments, and eight degrees of freedom, which was confirmed in AMOS. In the initial model, cEXC23 and cAPP were randomly chosen to have the scaling metric set to 1. After estimation, the indicators with the highest

unstandardized factor loadings had the scaling metric set to 1. All standardized regression weights were statistically significant and $\beta = .58$ or higher. Based on GFI and CFI model fit indices, the fit of the model was adequate; however, TLI and RMSEA statistics identified poor model fit (see Figure 10, Table 11, Model 1, $\chi^2 = 75.11$, $p = .000$, df = 8, GFI = .943, CFI = .936, TLI = .881, RMSEA = .136).

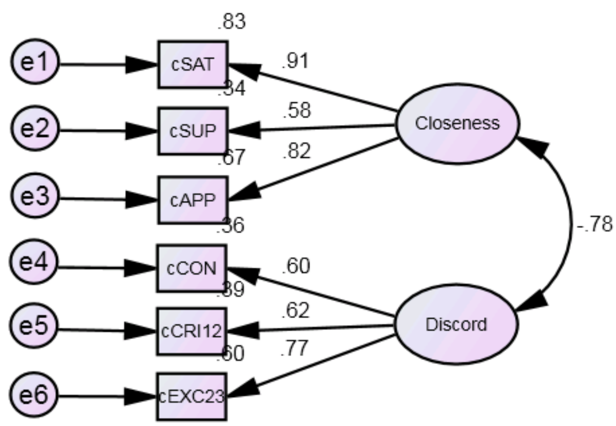


Figure 10. **NRI - Initial measurement model results**

Inspection of modification indices identified a possible improvement in the model if the error term for cSUP (e2) was allowed to covary with the error term for cAPP (e3). Support and approval are similar concepts, and support by a teacher can be considered approval by a teacher. Therefore, the theory supports covarying the error terms of these two constructs with significant model improvement; however, TLI and RMSEA values were still not acceptable (Model 2, $\chi^2 = 39.65$, $p = .000$, df = 7, GFI = .970, CFI = .969, TLI = .934, RMSEA = .101). Modification indices were again inspected, and the error terms for cCON (e4) and cCRI12 (e5) were identified to improve fit of the model (see Figure 11).

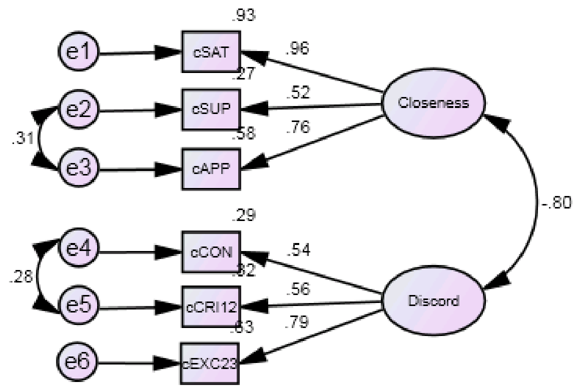


Figure 11. **Closeness and Discord results**

Conflict and criticism are related constructs and can be perceived similarly by students and theoretically can be set to covary. Model fit improved; all fit indices, including chi-square, indicated an excellent fitted model with all fit indices using the maximum-likelihood estimation method (Model 3, $\chi^2 = 12.20$, $p = .058$*, df = 6, GFI = .991, CFI = .994, TLI = .985, RMSEA = .048). All factor loadings and covariances were statistically significant with all standardized regression weights greater than $\beta = .52$. The standardized residual covariance matrix did not identify any indicators that did not fit the data.

Table 11

NRI Measurement model validation - Model fit indices utilizing composites

Model	χ^2	df	GFI	CFI	TLI	RMSEA	Notes
1	75.11	8	.943	.936	.881	.136	Closeness/Discord Composites
2	39.65	7	.970	.969	.934	.101	e2 and e3 set to covary (cSUP & cAPP)

3	12.20*	6	.991	.994	.985	.048	e4 and e5 set to covary (cCON & cCRI12)

* χ^2 not statistically significant

With the removal of cDIS, cPRE, iEXC1, and iCON3, the remaining composites had skewness values within |2| and kurtosis values within |7|, showing the univariate items were within accepted ranges of normal and could be analyzed utilizing the maximum-likelihood estimation method. According to Byrne (2014), multivariate kurtosis can be detrimental to SEM and CFA analysis. The multivariate critical ratio, which identifies multivariate normality (11.23), indicated the dataset was slightly multivariate non-normal. To confirm estimates calculated using the maximum-likelihood estimation method, the final model was confirmed using the Bayesian estimation method with nearly identical findings (see Table 12).

Table 12

NRI Regression weight comparisons between Maximum Likelihood and Bayesian estimates

			Maximum-Likelihood Estimation		Bayesian Estimation*	
Indicator		Latent Variable	Regression Weight	Std Error	Regression Weight	Std Error
cSAT	<---	Closeness	1		1	
cSUP	<---	Closeness	.438	.040	.439	.040

Table 12 Continued

			Maximum-Likelihood Estimation		Bayesian	

Table 12

NRI Regression weight comparisons between Maximum Likelihood and Bayesian estimates

					Estimation*	
Indicator		Latent Variable	Regression Weight	Std Error	Regression Weight	Std Error
cAPP	<---	Closeness	.740	.046	.745	.048
cCON	<---	Discord	.535	.054	.537	.057
cCRI12	<---	Discord	.586	.057	.586	.061
cEXC23	<---	Discord	1		1	

*Convergence statistic = 1.0011, 56,501 + 500 samples

Validation of measure - Confirmatory factor analysis – BPNS. A first-order CFA analysis was conducted using the nine items which were allowed to load on their corresponding constructs, with three items each for autonomy, competence, and relatedness. In the initial model, the scaling metric was set to one for iAUT1, iCOM1, and iREL1, with the scaling metric set to 1 on indicators with the highest unstandardized factor loadings after initial estimation (see Figure 12). Changes were made one at a time with the estimation process rerun between changes. Initial model estimates identified poor model fit through factor loadings and model fit indices (see Table 13, Model 1, χ^2 = 169.4, p = .000, df = 24, GFI = .920, CFI = .930, TLI = .895, RMSEA = .116).

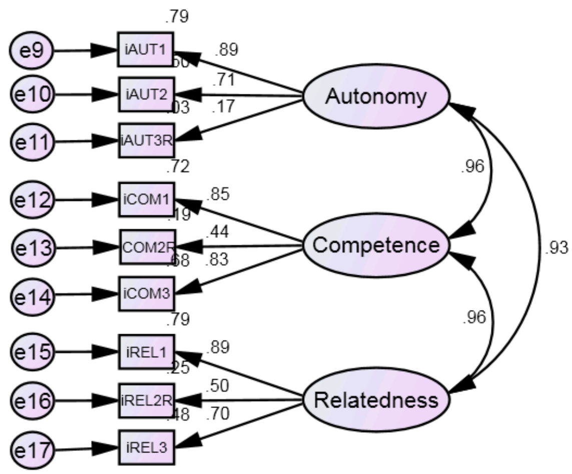


Figure 12. **BPNS - Initial measure validation results**

Based on descriptive statistics, student responses on the reverse worded items were atypical compared to the items that were not worded in the reverse direction. iAUT3R had a mean much higher than iAUT1 and iAUT2 and was more negatively skewed. The distribution of responses was also leptokurtic, while the other two were platykurtic. CFA results identified that iAUT3R factor loading was statistically significant ($\beta = .17$) and extremely low, with only 2.9% of the variance in iAUT3R explained by autonomy. The factor was dropped from the model due to the low loading and students struggling to respond to the question (Model 2, $\chi^2 = 90.1$, $p = .000$, df = 17, GFI = .952, CFI = .964, TLI = .940, RMSEA = .097). Model fit was good according to GFI and CFI fit indices, yet TLI and RMSEA indicated poor model specification.

iCOM2R suffered problems similar to iAUT3R. The mean was elevated well above iCOM1 and iCOM3, and was more negatively skewed. iCOM2R was also leptokurtic, while the other two indicators were platykurtic. The factor loading for iCOM2R was $\beta = .43$, with 18.9% of the variance explained by competence. The standardized residual covariance between iCOM2R and iREL2R (3.529) also indicated a problem. While the factor was statistically significant, similar to iAUT3R, the factor was

dropped from the model, which resulted in improved model fit; however, RMSEA was still high (Model 3, $\chi^2 = 56.41$, $p = .000$, df = 11, GFI = .963, CFI = .976, TLI = .954, RMSEA = .095).

The difference of iREL2R from the other indicators of relatedness was not as drastic as iAUT3R and iCOM2R; however, was still different. Responses to iREL2R were higher, more negatively skewed, and more peaked than the other two. The factor loading for iREL2R was $\beta = .50$ with 24.5% of the variance explained by relatedness, which was largely different from iREL1 and iREL3, and was removed from the model. GFI, CFI, and TLI indices indicated good model fit; however, RMSEA worsened and indicated adequate fit (see Figure 13, Table X, Model 4, $\chi^2 = 32.71$, $p = .000$, df = 6, GFI = .977, CFI = .985, TLI = .962, RMSEA = .099).

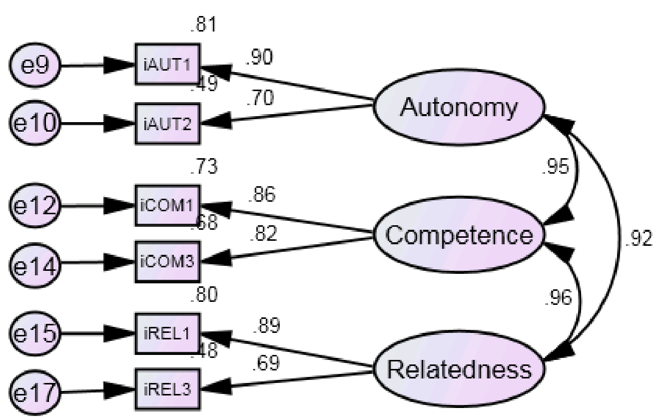


Figure 13. **BPNS - Final measure validation results**

The remaining items were statistically significant, had standardized regression weights greater than $\beta = .6$, and no issues were identified in the standardized residual covariance matrix. The six items of basic psychological needs satisfaction were used to calculate Cronbach's alpha (.91) and had a high reliability.

Table 13

BPNS Measure validation - Model fit indices

Model	χ^2	df	GFI	CFI	TLI	RMSEA	Notes
1	169.39	24	.920	.930	.895	.116	Original Model
2	90.06	17	.952	.964	.940	.097	iAUT3R removed
3	56.41	11	.963	.976	.954	.095	iCOM2R removed
4	32.71	6	.977	.985	.962	.099	iREL2 removed

Validation of measurement model - Confirmatory factor analysis – BPNS. The first-order measurement model of basic psychological needs consisted of six questions, as the three reverse worded questions were removed. The model had 15 free parameters to be estimated, 21 sample moments, and six degrees of freedom, which was confirmed in AMOS.

Modification indices indicated better model fit if the error term for iAUT2 (e10) and iCOM1 (e12) were set to covary (see Figure 14). iAUT1 asked, “When I am with my teacher, I have a say in what happens and I can voice my opinion” while iCOM1 asked, “When I am with my teacher, I feel like a competent person.” A student that has a say and can voice his or her opinion in a classroom could be expected to feel competent and, therefore, have similar responses to both questions (Tian et al., 2014).

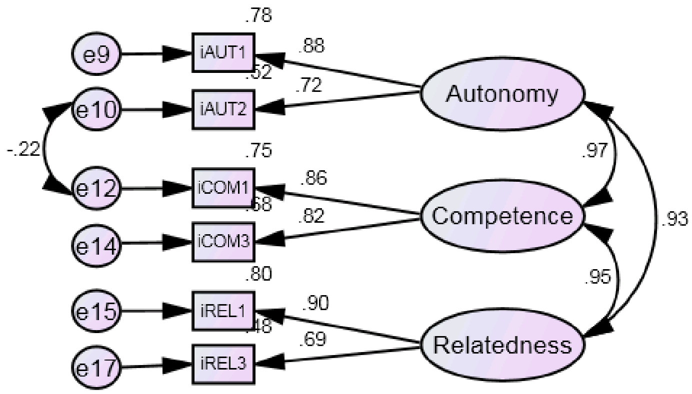


Figure 14. **BPNS - Final measurement model results**

The questions/constructs are interrelated, and based on prior research findings, theoretically, the error terms could be allowed to covary. Model fit statistics improved; however, RMSEA indicated a poorly fitted model using maximum-likelihood estimation (see Table 14, χ^2 = 20.69, p = .001, df = 5, GFI = .986, CFI = .991, TLI = .973, RMSEA = .083.

Table 14

BPNS Measurement model validation - Model fit indices

Model	χ^2	df	GFI	CFI	TLI	RMSEA	Notes
1	32.71	6	.977	.985	.962	.099	Original Model
2	20.69	5	.986	.991	.973	.083	e10 and e12 set to covary (iAUT2 & iCOM1)

With the removal of the reverse worded questions iAUT3R, iCOM2R, and iREL2R, the remaining items had skewness values within |1| and kurtosis values within |1.3|, showing the items were normal and could be analyzed using the maximum-likelihood estimation method. The multivariate critical ratio, which identifies multivariate normality (8.957), indicated the dataset was very close to multivariate

normal. To confirm maximum-likelihood estimations, the final model was confirmed using the Bayesian estimation method with nearly identical findings (see Table 15).

Table 15

BPNS Regression weight comparisons between Maximum Likelihood and Bayesian estimates

			Maximum-Likelihood Estimation		Bayesian Estimation*	
Indicator		Latent Variable	Regression Weight	Std Error	Regression Weight	Std Error
iCOM3	<---	Competence	0.905	.041	.905	.042
iCOM1	<---	Competence	1			
iREL3	<---	Relatedness	.796	.047	.795	.048
iREL1	<---	Relatedness	1			
iAUT2	<---	Autonomy	.829	.048	.828	.049
iAUT1	<---	Autonomy	1			

*Convergence statistic = 1.0015, 71,501 + 500 samples

Validation of measure - Confirmatory factor analysis – CEI. The CEI used in this study was modified from the original version, using only thirteen of the original twenty-four items and measuring four dimensions of engagement of the original five to include affective engagement, behavioral engagement (compliance), behavioral engagement (effortful class participation) and cognitive engagement.

Wang et al. (2014) explored the factor structure of engagement when developing the CEI instrument, and their analysis indicated that a first-order multidimensional model

was most appropriate when analyzing the dimensions of engagement. In the initial model, all items were used to construct a first-order model, with individual items loading on the corresponding constructs of affective engagement, behavioral engagement compliance, behavioral engagement participation, and cognitive engagement (see Figure 15).

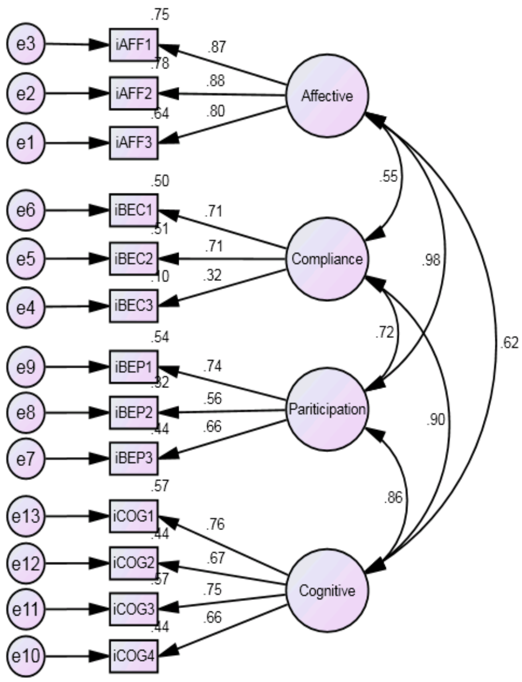


Figure 15. **CEI Initial measure validation results**

A scaling metric of 1 was set for the first question in each dimension of engagement, with the scaling metric set to 1 for indicators with the highest unstandardized factor loadings following estimation. The model was estimated, and changes were made one at a time with the estimation process run after every change (Model 1, χ^2 = 220.64, *p* = .000, df = 59, GFI = .930, CFI = .943, TLI = .924, RMSEA = .078).

AMOS reported the solution was not admissible because there was a non-positive definite covariance matrix. According to Wothke (1993), non-positive covariance

matrices can be caused by the presence of outliers and non-normal data, too many parameters, empirical under identification, and model misspecification. Rigdon (1997) stated non-positive definite matrices could also be the result of perfect linear dependency of one indicator variable on another, or several variables together perfectly predicting another variable (multicollinearity).

Univariate and multivariate outliers have been addressed. Non-normal data could be an issue because iBEC3 (sk = -2.456, k = 6.515) was highly negatively skewed and leptokurtic. There were 13 variables used to identify four latent variables, which when compared to many other CFA models, would be considered simple, with a satisfactory number of variables, and the model is identified. The highest correlation on the sample correlations table for indicator variables was between iAFF2 and iAFF1 (r = .764), which was expected as the questions covered common content. All other correlations were lower, so perfect linear dependency of one indicator variable on another was not an issue.

The implied correlation between the latent variables affective engagement and behavioral engagement participation (r = .99), and between behavioral engagement compliance and cognitive engagement (r = .90), were high. Individually, the items of the affective engagement dimension and behavioral engagement participation dimension and cognitive engagement and behavioral engagement compliance are not highly correlated; however, the combined effects of the items are. Inspection of the implied correlation matrix identified affective engagement items having a high correlation with both the behavioral engagement participation (r_{iAFF1} = .854, r_{iAFF2} = .868, r_{iAFF3} = .788) and affective engagement factors (r_{iAFF1} = .867, r_{iAFF2} = .881, r_{iAFF3} = .800). Due to the high correlation between latent variables and lower correlations between indicator items, it

appeared that several indicator variables together almost perfectly predicted another variable and may have caused the non-positive definite matrix.

Wothke (1993) stated non-positive definite matrices can sometimes be rectified by removing items from the model. iBEC3, was highly negatively skewed, leptokurtic, did not mirror question iBEC2 (*sk* = -1.773, *k* = 2.532) and iBEC1 (*sk* = -1.812, *k* = 2.697), and had a standardized regression weight of β = .32 in the first model, and was removed from the analysis. The model fit indices improved slightly but was still inadmissible, with a non-positive definite matrix (see Table 16, Model 2, χ^2 = 185.42, *p* = .000, df = 48, GFI = .936, CFI = .950, TLI = .931, RMSEA = .079). The implied correlation between the latent variables affective engagement and behavioral engagement participation remained the same (*r* = .99) and increased between behavioral engagement compliance and cognitive engagement (*r* = .91).

Similar to iBEC3, iBEP2 (*M* = 2.98, *sk* = -.057, *k* = -1.449) had a question that was different from the other items in the construct, and resulted in a lower mean and skewness, and was more platykurtic compared to iBEP1 (*M* = 3.59, *sk* = -.570, *k* = -1.061) and iBEP3 (*M* = 4.02, *sk* = -.938, *k* = -.119). The observed standardized regression weight of iBEP2 (β = .57), was lower than iBEP1 (β = .74) and iBEP3 (β = .66), and was removed from the model (Model 3, χ^2 = 166.98, *p* = .000, df = 38, GFI = .937, CFI = .950, TLI = .927, RMSEA = .86). Model fit was slightly worse; however, the correlation between affective engagement and behavioral engagement participation (r = .96) decreased. The solution was still not admissible. In an attempt to address the high correlation between affective engagement and behavioral engagement participation, iAFF3 was removed from the model since it had the lowest loading (β = .79) compared to

iAFF1 (β = .87) and iAFF2 (β = .89). Model fit indices indicated an adequate to good fitting model, and the solution was now admissible (Model 4, $\chi^2 = 99.18$, $p = .000$, df = 29, GFI = .959, CFI = .967, TLI = .949, RMSEA = .073). The correlation between affective engagement and behavioral engagement participation dropped ($r = .91$). With an admissible solution and no other items having been different from others in the respective construct, no other items were deleted.

Modification indices identified an improvement in the model if the error term for iBEC1 and iCOG1 were set to covary. The high correlation between the factor behavioral engagement compliance and cognitive engagement identified high content overlap, therefore, the terms were allowed to covary. The final model improved slightly and had good fit according to GFI, CFI, and TLI statistics and adequate fit according to RMSEA statistics (see Figure 16, Model 5, $\chi^2 = 79.47$, $p = .000$, df = 28, GFI = .966, CFI = .976, TLI = .962, RMSEA = .064).

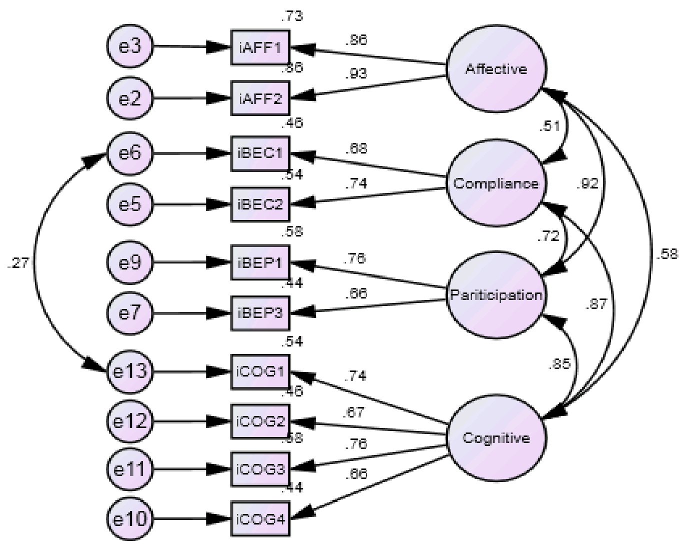


Figure 16. CEI - **Final measure validation results**

There were no warning signs in the standardized residual covariance matrix of model misspecification, and all remaining items had statistically significant estimates with standardized regression weights greater than .6. Cronbach's alpha for classroom engagement were calculated (.88, .66, .66, and .80) for affective engagement, behavioral engagement (compliance), behavioral engagement (effortful class participation) and cognitive engagement respectively, compared to Sever et al, (2014) findings of .87, .74, .82, and .89. To avoid generating an over-fitted model, no more modifications were made to the model.

Table 16

CEI Measure validation - Model fit indices

Model	χ^2	df	GFI	CFI	TLI	RMSEA	Notes
1	220.64	59	.930	.943	.924	.078	Original Model. Solution not admissible.
2	185.42	48	.936	.950	.931	.079	iBEC3 removed. Solution not admissible
3	166.98	38	.937	.950	.927	.086	iBEP2 removed. Solution not admissible.
4	99.18	29	.959	.967	.949	.073	iAFF3 removed. Solution admissible.
5	79.47	28	.966	.976	.962	.064	e6 and e13 set to covary (iBEC1 & iCOG1)

Validation of measurement model - Confirmatory factor analysis – CEI. The remaining indicator items from measurement validation were used to build parcels. Parcels were computed for affective engagement (cAFF12 included iAFF1 and iAFF2), behavioral engagement compliance (cBEC12 included iBEC1 and iBEC2), behavioral

engagement participation (cBEP13 included iBEP1 and iBEP3), and cognitive engagement (cCOG included iCOG1, iCOG2, iOCG3, and iCOG4) by adding the scores of the individual items and dividing by the number of items in the respective dimension. In the initial model, cAFF12 was randomly chosen and assigned a scaling metric of one, which was moved to cBEP13 after initial estimation (see Figure 17).

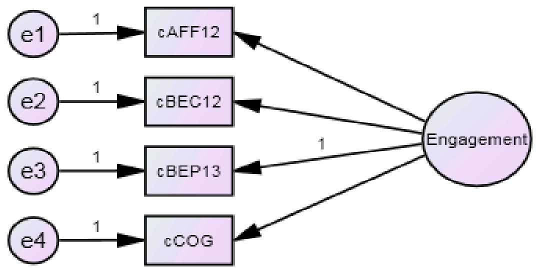


Figure 17. **CEI - Initial measurement model**

Initial model fit estimates showed poor model fit (Model 1, χ^2 = 115.55, p = .000, df = 2, GFI = .883, CFI = .862, TLI = .585, RMSEA = .354); however, all indicator variables were statistically significant, and all standardized loadings were greater than β = .5 (see Table 17). Model fit was improved by setting the error term for cBEC12 (e2) to covary with the error term for cCOG (e4, see Figure 18). The dimension of behavioral engagement compliance and cognitive engagement had a high correlation in earlier examination, so error terms should be highly correlated and, therefore, allowed to covary (Model 2, χ^2 = .248, p = .618*, df = 1, GFI = 1.00, CFI = 1.00, TLI = 1.00, RMSEA = .000). Model fit was perfect according to model fit indices; all loadings were statistically significant; and there were no warning signs in the standardized residual matrix.

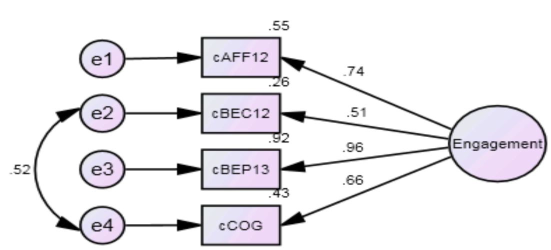


Figure 18. **CEI - Final measurement model results**

Table 17

CEI Measurement model validation - Model fit indices

Model	χ^2	df	GFI	CFI	TLI	RMSEA	Notes
1	115.55	2	.883	.862	.585	.354	Measurement model with composites
2	.248*	1	1.00	1.00	1.00	.000	e2 and e4 set to covary (cBEC12 & cCOG)

* χ^2 not statistically significant

While all parceled indicators had skewness and kurtosis values with |2| and |7| respectively, similar to the measurement models for NRI and BPNS, CEI was confirmed using the Bayesian estimation method with results nearly identical to maximum-likelihood estimates (see Table 18).

Table 18

CEI Regression weight comparisons between Maximum Likelihood and Bayesian estimates

			Maximum Likelihood Estimation		Bayesian Estimation*	
Indicator		Latent Variable	Regression Weight	Std Error	Regression Weight	Std Error
cAFF12	<---	Engagement	.897	.060	.898	.062
cBEC12	<---	Engagement	.371	.036	.371	.037
cBEP13	<---	Engagement	1		1	

cCOG	<---	Engagement	.537	.040	.538	.041

*Convergence statistic = 1.0011, 73,501 + 500 samples

Validation of measurement model - Confirmatory factor analysis – Outcome. Outcome was measured using teacher generated GPA's and fourth term averages (Trm4Avg), along with state generated scale scores (ScScr) and norm-referenced scores (NormRef) on the Georgia Milestones assessment. Confirmatory factor analysis for a first order model was conducted with the scaling metric set to 1 on ScScr after initial calculation of estimates (see Figure 19).

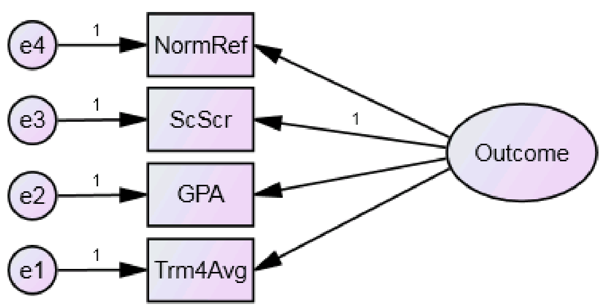


Figure 19. **Outcome - Initial measurement model**

Model fit indices indicated a poorly fit model; however, all factor loadings were statistically significant and in the proper direction (see Table 19, Model 1, $\chi^2 = 238.29$, $p = .000$, df = 2, GFI = .825, CFI = .812, TLI = .435, RMSEA = .510). Inspection of modification indices identified the error terms of ScScr and NormRef to be set to covary. The standardized residual covariance between the two was 8.589, which also identified a problem with model specification. These two measures are both state-generated and were highly correlated, so covarying these terms fell in the realm of possibility. Model fit was exceptional, and all indicators were statistically significant (see Table 19, Model

2, χ^2 = 1.394, p = .238*, df = 1, GFI = .998, CFI = 1.0, TLI = .998, RMSEA = .029) with a reliability coefficient of .69. The standardized regression weight of GPA was β = .99, with 100% of the variance in GPA explained by outcome.

Table 19

Outcome measurement model validation - Model fit indices

Model	χ^2	df	GFI	CFI	TLI	RMSEA	Notes
1	238.29	2	.825	.812	.435	.510	Original Model.
2	1.39*	1	.998	1.00	.998	.029	e3 and e4 set to covary (ScScr & NormRef)

* χ^2 not statistically significant

Structural Equation Modeling

The following five steps were used to conduct the SEM analysis: model specification, identification, estimation, testing, and modification.

Model specification. Based on prior research and validation of the measuring instruments, the following structural equation model was hypothesized based on the Self-Systems Process Model and was linear in nature (see Figure 20).

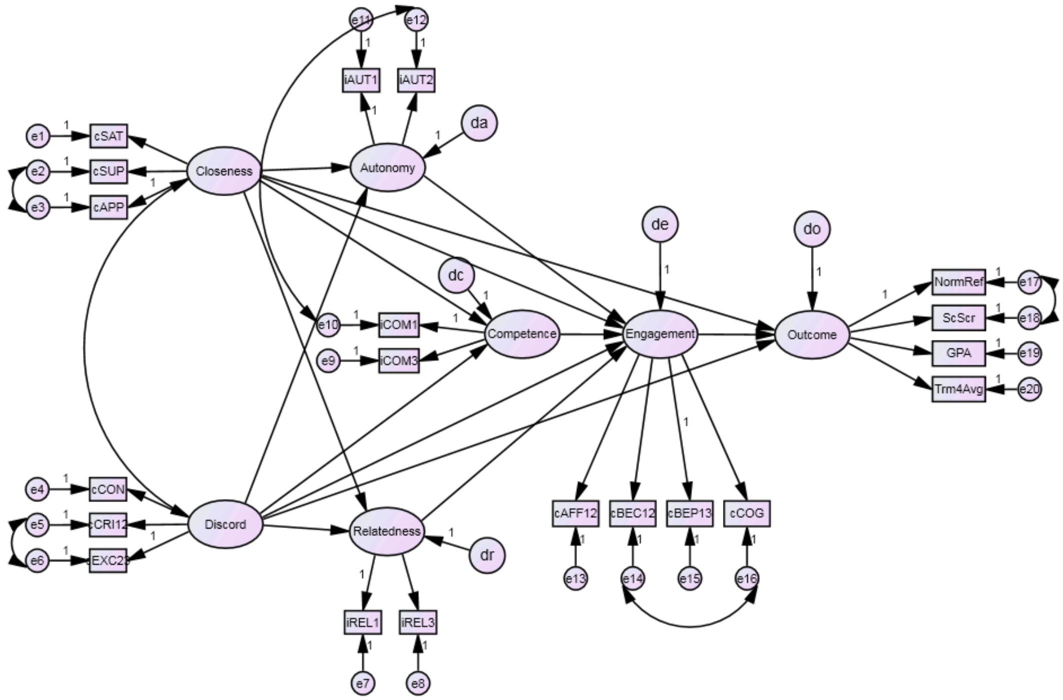


Figure 20. **Hypothesized structural model of the impact of NRI on BPNS, CEI, and student Outcomes.**

There were seven latent variables in the proposed model, which were represented by ovals, and they include closeness, discord, autonomy, competence, relatedness, engagement, and outcome. Each construct was measured by the corresponding indicators previously validated and also included covariances that were determined from measurement models. Composites were created for closeness, discord, and engagement rather than individual items to simplify the hypothesized model, to create a more continuous measurement scale, and to alleviate the issue of needing an unobtainable number of student responses.

Factorial analysis of the NRI identified closeness and discord as second order factors (Furman & Buhrmester, 2009). Closeness was comprised of three first order factors that consisted of satisfaction (cSAT), support (cSUP), and approval (cAPP) after the removal of disclosure, each being assessed by three items. Discord was comprised of

three first order factors that consisted of conflict (cCON), criticism (cCRI12), and exclusion (cEXC23), each being assessed by two items except conflict being assessed by three. Parcels were created by adding the items for each factor and dividing by the number of items used to generate the parcel.

The basic psychological needs inventory originally contained nine items, with three items each for autonomy, competence, and relatedness. Due to issues with the negatively worded questions, one question for each autonomy, competence, and relatedness was removed during CFA and the negatively worded questions were not included in the structural model. Two items were allowed to load on each factor, and the items were not parceled. Parcels created for engagement consisted of two items for affective engagement (cAFF2), two items for behavioral engagement compliance (cBEC12), two items for behavioral engagement effortful class participation (cBEP13), and four items for cognitive engagement (cCOG). The four indicators of outcome included GPA, Term4Avg, NormRef, and ScScr and were not parceled. Closeness and discord, measures of the teacher-student relationship (context), were hypothesized to affect autonomy, competence, and relatedness (self), which all influence engagement (action), which consequently influences NormRef, ScScr, GPA, and Term4Avg (outcome).

Model identification. In order for AMOS to estimate a unique value for every parameter in the model, the model must be identified and have a degrees of freedom value greater than 0 (Kline, 2011). All 20 indicators had an associated error term (circle) which identified that each indicator had non-random or measurement error (Byrne, 2010). The five latent endogenous variables of autonomy (Da), competence (Dc), relatedness

(Dr), engagement (De), and outcome (Do) had a corresponding disturbance term. Closeness and discord had an associated variance term and were also set to covary. There were 13 factor loadings, 14 path coefficients, 20 error variances, 5 disturbances, 2 variances, and 6 covariances for a total of 60 free parameters that were estimated. Degrees of freedom is a function of the number of observed variables in the model (Hoyle, 2014). Using the equation $p (p + 1) / 2$, where $p = 20$ and was the number of observed parameters in the hypothesized model, there were 210 elements in the correlation matrix. Degrees of freedom, the difference between known and unknown information (Hoyle, 2014), was determined to be $df = 150$ by subtracting the number of free parameters to be estimated, 60, from the number of elements in the correlation matrix, 210, which was confirmed by AMOS.

Model estimation.

Sample size. SEM is a large sample technique requiring a large number of responses with a minimum sample size of 200 (Kline, 2011). Teo, Tsai, & Yang, (2013) and In’nami & Koizumi (2013) recommended that the sample size be equal to 10 participants per parameter estimated. If the observed data is not normal, sample size should be increased to 15 participants per parameter (Teo et al., 2013). The more complex the model, the larger the sample size is required (Kline, 2011; Teo et al., 2013). The number of parameters to be estimated based on the hypothesized model was 60. A sample size between 200 and 600 was recommended by the literature, with the latter being better in the situation that the model is complex and data non-normal.

Univariate and multivariate normality. According to Newsom, skewness values greater than |2| and kurtosis greater than |7| indicate a variable is non-normal, with

kurtosis values being more important than skewness. In'nami & Koizumi (2013) indicated that values exceeding |3| and |21| for skewness and kurtosis values respectively are extremely non-normal, while a skewness of |2| and a kurtosis of |7| is moderately non-normal. The data, while not perfectly normal, had skewness and kurtosis values less than |2| and |7| respectively allowing for a smaller number of participants. Maximum-likelihood estimation methods assume no excessive kurtosis of observed variables (Hoyle, 2015). Many of the items in the analysis had skewness and kurtosis values less than |1| and |2| respectively, however cBEC12 (sk = -1.644, k = 2.296), cCRI12 (sk = 1.559, k = 2.141), and cCON (sk = 1.797, k = 2.845) exhibited signs of being non-normal (see Table 20). Multivariate normality was assessed using Mardia's normalized estimate (16.181), which indicated modest multivariate non-normality. Descriptive statistics allowed for maximum-likelihood estimation methods; however, the alternate Bayesian estimation was used to verify findings.

Table 20

Univariate and Multivariate normality of SEM indicators

Variable	skewness	c.r.	kurtosis	c.r.
cSAT	-0.542	-4.723	-0.698	-3.041
cSUP	0.844	7.347	0.213	0.926
cAPP	-0.139	-1.208	-0.897	-3.905
cCON	1.797	15.651	2.845	12.389
cCRI12	1.559	13.574	2.141	9.321
cEXC23	1.21	10.538	0.694	3.023
iAUT2	-0.07	-0.608	-1.214	-5.285
iAUT1	-0.351	-3.059	-1.041	-4.532

iCOM1	-0.495	-4.311	-0.894	-3.893
iCOM3	-0.539	-4.692	-0.799	-3.479
iREL1	-0.51	-4.44	-0.95	-4.138
iREL3	0.257	2.234	-1.168	-5.085
cCOG	-1.049	-9.133	0.34	1.482
cAFF12	-0.511	-4.446	-0.955	-4.158
cBEC12	-1.644	-14.318	2.296	9.996
cBEP13	-0.687	-5.983	-0.526	-2.292
NormRef	-0.872	-7.593	0.013	0.058
ScScr	0.164	1.43	0.619	2.695
GPA	-0.758	-6.603	-0.081	-0.355
Trm4Avg	-0.758	-6.601	0.213	0.928
Multivariate			45.005	16.181

The estimation process was performed, and results were poor with unacceptable model fit indices (Model 1, χ^2 = 479.5, p = .000, df = 150, GFI =.903, CFI = .945, TLI = .930, RMSEA = .070), along with a negative variance for the disturbance on relatedness (dr = -.076). Relatedness had a squared multiple correlation greater than 1, there were five standardized regression weights greater than 1, and there were large standard errors on multiple latent constructs which indicated a problem with the structural model. AMOS reported the solution was inadmissible due to the negative variance. The many issues identified in the model indicated possible multicollinearity in the model.

Multicollinearity. Multicollinearity may result when two independent indicators or factors have intercorrelations greater than .9 (Byrne, 2010; Kline, 2011; Tabachnick & Fidell, 2013), a variance inflation factor (VIF) greater than 4, and/or a tolerance less than .20 (Garson, 2012). The tolerance statistic takes into account the interaction effect of

other independent variables as well as correlations between the variables (Garson). Multicollinearity can be adjusted for by deleting or combining redundant indicators or constructs (Kline; In'nami & Koizumi, 2013; Tabachnick & Fidell, 2013). Following the validation of each measurement model, factor scores were imputed into SPSS for each latent variable. SPSS was used to compute bivariate correlations between all indicators and between the latent constructs of closeness, discord, autonomy, competence, relatedness, and engagement because these were the independent variables used to predict outcome, the dependent variable.

Sample correlations between all indicators in the model were below $r = .9$. The highest correlation ($r = .884$) occurred between Term4Avg and GPA and was the result of the indicators being so closely related. The variance inflation factor score ($VIF = 4.56$) and tolerance statistic ($Tol = .219$) for GPA and Term4Avg, along with the high correlation indicated multicollinearity to be present, and Term4Avg was dropped from the model. The reliability coefficient of NormRef, ScScr, and GPA was recalculated with nearly identical results (Cronbach's Alpha = .678).

The latent variables of closeness and discord had an elevated correlation ($r = -.888$), which was expected as closeness is essentially the opposite in meaning. The correlations among the factors of autonomy and competence ($r = .995$), autonomy and relatedness ($r = .980$), and competence and relatedness ($r = .989$) were exceptionally high and indicated multicollinearity to be a threat to the structural equation model (see Table 21).

Table 21

Pearson bivariate correlations between latent constructs

	Closeness	Discord	Auto	Related	Comp	Engage	Outcome
Closeness	1						
Discord	-.888**	1					
Autonomy	.764**	-.709**	1				
Relatedness	.792**	-.719**	.980**	1			
Competence	.771**	-.710**	.995**	.989**	1		
Engagement	.593**	-.542**	.659**	.645**	.653**	1	
Outcome	.174**	-.169**	.204**	.186**	.207**	.196**	1

** Correlation is significant at the 0.01 level (2-tailed).

Variance inflation factor and tolerance scores were also computed for the latent constructs of closeness, discord, autonomy, competence, relatedness, and engagement with outcome as the dependent variable. Autonomy, competence, and relatedness had VIF scores much greater than 4 and Tolerance scores lower than .2 (see Table 22).

Table 22
Collinearity statistics between latent constructs

	Tolerance	VIF
Closeness	0.154	6.474

Discord	0.206	4.854
Autonomy	0.008	122.085
Relatedness	0.018	56.202
Competence	0.005	219.252
Engagement	0.544	1.84

a Dependent Variable: Outcome

The extreme violation of correlation, tolerance, and VIF scores indicated multicollinearity was an issue with the original hypothesized model. According to Kline (2011); In'nami & Koizumi (2013); Tabachnick & Fidell (2013), multicollinearity can be adjusted for by deleting or combining redundant variables. Autonomy, competence, and relatedness were all indicators of the same construct and were collapsed into a single latent variable termed basic psychological needs (BPNS), with 6 indicator variables, 2 indicator variables from each original latent variables (see Figure 21).

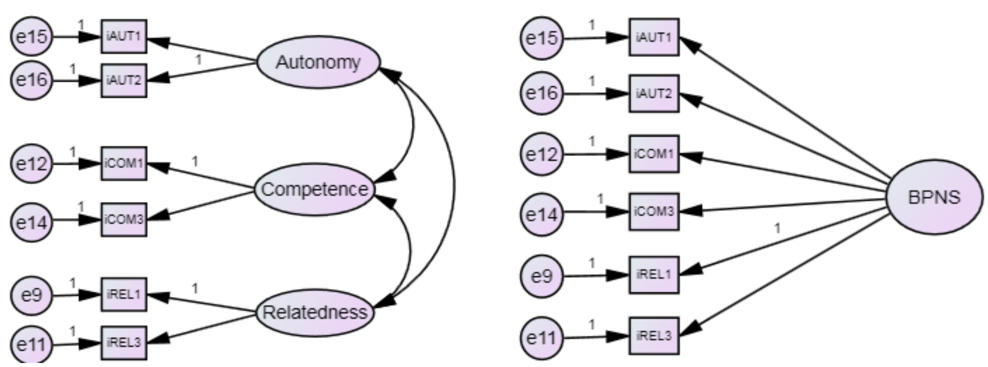


Figure 21. **BPNS - Collapsed construct measurement model**

Confirmatory factor analysis was conducted on the new model with the scaling metric set to 1 on iAUT1. All indicators were statistically significant and greater than .6, with model fit indices showing adequate model fit (see Table 23, Model 1, $\chi^2 = 44.47$, *p*

= .000, df = 9, GFI =.971, CFI = .980, TLI = .966, RMSEA = .093). There error terms iAUT2 and iCOM1 (Model 2, χ^2 = 30.27, p = .000, df = 8, GFI =.979, CFI = .987, TLI = .976, RMSEA = .078) and iREL3 and iCOM1 (Model 3, χ^2 = 12.14, p = .096*, df = 7, GFI =.992, CFI = .997, TLI = .994, RMSEA = .040) were set to covary to improve model fit based on modification indices (see Figure 22).

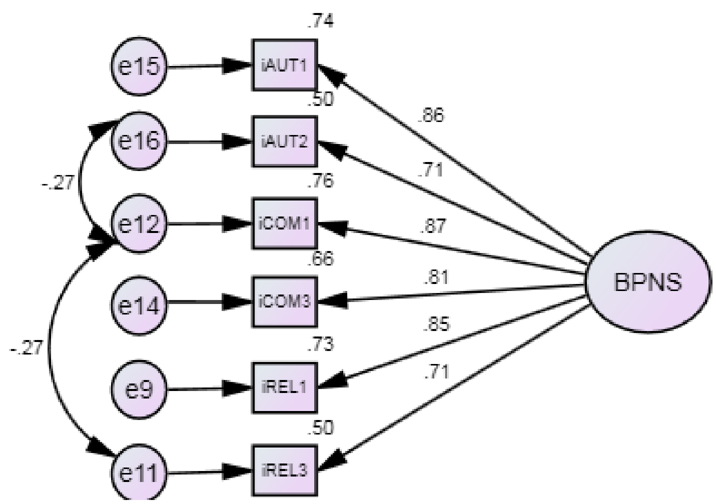


Figure 22. BPNS - Collapsed construct measurement model results

In the final model, all regression weights were statistically significant and standardized regressions weights were greater than .7 with excellent model fit indices.

Table 23

BPNS - Collapsed measurement model validation - Model fit indices

Model	χ^2	df	GFI	CFI	TLI	RMSEA	Notes
1	44.47	9	.971	.980	.966	.093	Original Model
2	30.27	8	.979	.987	.976	.078	e12 and e16 set to covary (iCOM1 & iAUT2)

Table 23

BPNS - Collapsed measurement model validation - Model fit indices

3	12.14*	7	.992	.997	.994	.040	e11 and e12 set to covary (iREL3 & iCOM1)

* χ^2 not statistically significant

Factor scores for BPNS were imputed into SPSS, and Pearson bivariate correlations were again calculated between latent constructs (see Table 24). There were no correlations greater than .9; however, closeness and discord were high (*r* = -.89).

Table 24

Pearson bivariate correlations between latent constructs - Retest 1

	Closeness	Discord	BPNS	Engagement	Outcome
Closeness	1				
Discord	-.888**	1			
BPNS	.772**	-.706**	1		
Engagement	.593**	-.542**	.652**	1	
Outcome	.174**	-.169**	.200**	.196**	1

** Correlation is significant at the 0.01 level (2-tailed).

Closeness and discord continued to have VIF scores greater than 4 and tolerance values around .2, which indicated multicollinearity between these two variables (see Table 25).

Table 25

Collinearity statistics between latent constructs - Retest 1

	Tolerance	VIF
Closeness	0.168	5.947
Discord	0.211	4.737
BPNS	0.345	2.9
Engagement	0.555	1.801

a Dependent Variable: Outcome

The constructs of closeness and discord were combined into a single latent variable (Kline, 2013; In'nami & Koizumi, 2013; Tabachnick & Fidell, 2013) teacher-student relationship (TSR), with 6 indicator variables; three each from closeness and discord (see Figure 23).

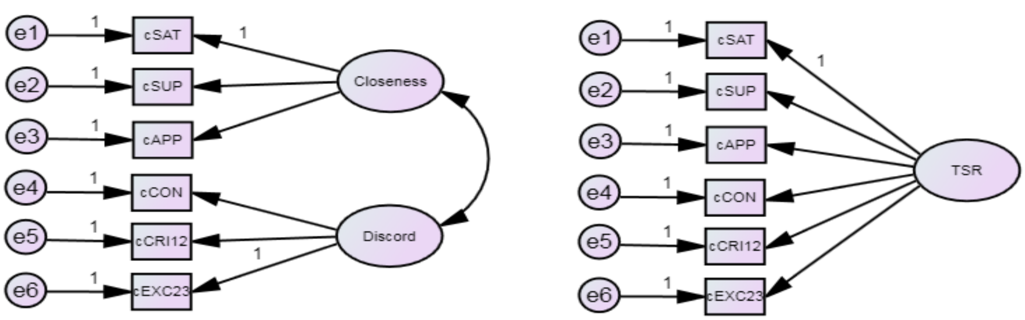


Figure 23. **TSR - Collapsed construct measurement model results**

The collapsed model was analyzed with the scaling metric set to 1 on cSAT, which ended up having the highest unstandardized regression weight. The regression weights of all indicators were statistically significant and in the direction expected, with factors of closeness being positive and factors of discord being negative; however, model

fit was poor (see Table 26, Model 1, χ^2 = 140.52, *p* = .000, df = 9, GFI =.891, CFI = .875, TLI = .792, RMSEA = .179). Modification indices identified the error terms for cCON and cCRI12 (Model 2, χ^2 = 77.38, *p* = .000, df = 8, GFI =.941, CFI = .934, TLI = .877, RMSEA = .138), cSUP and cAPP (Model 3, χ^2 = 43.65, *p* = .000, df = 7, GFI =.968, CFI = .965, TLI = .926, RMSEA = .107), cCON and cEXC23 (Model 4, χ^2 = 31.74, *p* = .039, df = 6, GFI =.977, CFI = .976, TLI = .939, RMSEA = .097), and cCRI12 and cEXC23 (Model 5, χ^2 = 11.58, *p* = .041, df = 5, GFI =.992, CFI = .994, TLI = .982, RMSEA = .054) to improve model fit if set to covary.

Table 26

TSR Measurement model validation - Model fit indices

Model	χ^2	df	GFI	CFI	TLI	RMSEA	Notes
1	140.52	9	.891	.875	.792	.179	Original Model
2	77.38	8	.941	.934	.877	.138	e4 and e5 set to covary (cCON & cCRI12)
3	43.65	7	.968	.965	.926	.107	e2 and e3 set to covary (cSUP & cAPP)
4	31.74	6	.977	.976	.939	.097	e4 and e6 set to covary (cCON & cEXC23)
5	11.73	5	.992	.994	.981	.054	e5 and e6 set to covary (cCRI12 & cEXC23)

GFI, CFI, and TLI values indicated an excellent fitted model, while RMSEA indicated adequate model fit. All regression weights were statistically significant with standardized regressions weights greater than .4 and in the proper direction with measures of discord negative (see Figure 24).

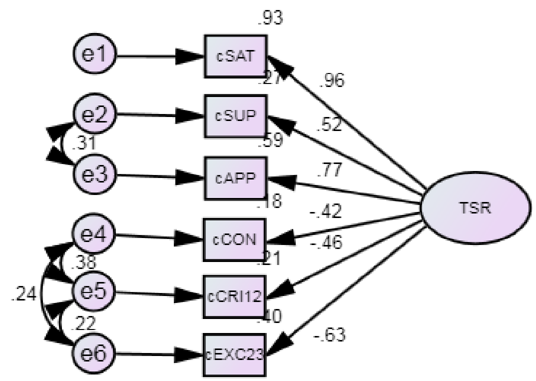


Figure 24. **TSR - Collapsed construct measurement model results**

Factor scores for TSR were imputed into SPSS, and Pearson bivariate correlations were calculated between latent constructs (see Table 27).

Table 27

Pearson bivariate correlations between latent constructs - Retest 2

	TSR	BPNS	Engagement
BPNS	.772**	1	
Engagement	.593**	.652**	1
Outcome	.174**	.200**	.196**

** Correlation is significant at the 0.01 level (2-tailed).

There highest correlation was between TSR and BPNS (r = .772), and collinearity statistics no longer identified multicollinearity to be an issue (see Table 28, *VIF* = 2.567, *Tol* = .39).

Table 28

Collinearity statistics between latent construct - Retest 2

Construct	Tolerance	VIF
TSR	0.39	2.567

Table 28 Continued

Construct	Tolerance	VIF
BPNS	0.346	2.89
Engagement	0.555	1.801

a Dependent Variable: Outcome

The measurement model for outcome was modified by removing the indicator Term4Avg. With the removal of Term4Avg, the construct of outcome contained only three indicators and CFA could not be conducted, and the covariance between the error terms of ScScr and NormRef, determined from the previous CFA, was removed. Closeness and discord was replaced with TSR, while autonomy, competence, and relatedness was replaced with BPNS. Scaling metrics were set to 1 for the indicators of cSAT, iAUT1, cAFF12 and ScScr after initial estimation (see Figure 25).

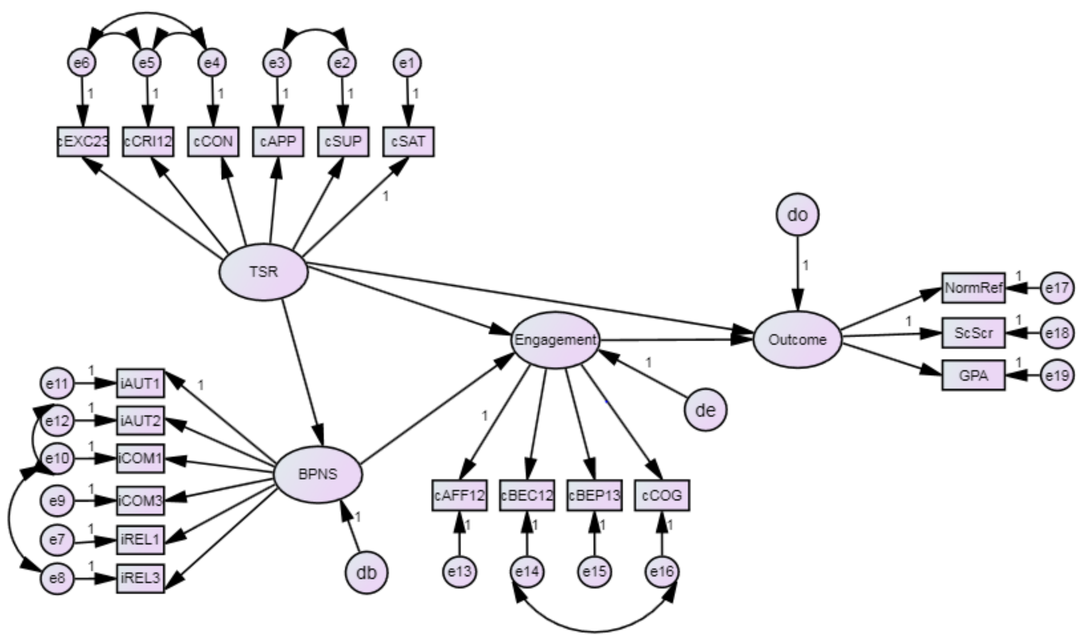


Figure 25. **Modified structural model based on measurement model validation.**

Scale score structural model testing. In the new structural model, there were 15 factor loadings, 5 path coefficients, 19 error variances, 3 disturbances, 1 variance and 7 covariances for a total of 50 free parameters that were estimated. Using the equation $p\ (\ p + 1\ )\ /\ 2$, where $p = 19$ and was the number of observed parameters in the hypothesized model, there were 190 elements in the correlation matrix. Degrees of freedom was determined to be $df = 140$ by subtracting the number of free parameters to be estimated 50 from the number of elements in the correlation matrix 190, which was confirmed by AMOS. Following estimation, the negative variance causing the inadmissible solution was no longer present. Model fit was poor (Model 1, $\chi^2 = 447.80$, $p = .000$, df = 140, GFI =.905, CFI = .942, TLI = .929, RMSEA = .070), with all paths except TSR on outcome statistically significant (see Table 29).

Table 29

Full structural model - Model fit indices

Model	χ^2	df	GFI	CFI	TLI	RMSEA	Notes
1	447.80	140	.905	.942	.929	.070	Full model with errors terms set to covary determined in measurement models
2	419.45	139	.911	.947	.935	.067	e15 and e16 set to covary (cBEP13 & cCOG)
3	404.72	138	.915	.949	.937	.065	e7 and e12 set to covary (iREL1 & iAUT2)
4	300.68	122	.933	.964	.955	.057	cSUP removed from the model
5	300.75	123	.933	.964	.956	.056	Nonsignificant path from TSR to outcome removed

Standardized regression weights were negative for aspects of discord, which was expected. The strongest relationship existed between TSR and BPN (β = .874), while the weakest relationship existed between TSR and outcome (β = .022, p = .820). Modification indices, standardized residual covariances, and model fit indices were used to identify the acceptability of the model. The error terms for cBEP13 and cCOG (Model 2, χ^2 = 419.45, p = .000, df = 139, GFI =.911, CFI = .947, TLI = .935, RMSEA = .067) and iAUT2 and iREL1 (Model 3, χ^2 = 404.72, p = .000, df = 138, GFI =.915, CFI = .949, TLI = .937, RMSEA = .065) were set to covary using modification indices as a guide.

Modification indices of regression weights showed significant improvement in the model if outcome, all the factors of outcome, and iREL3 were allowed to load on cSUP, which was in the wrong direction and not hypothesized, and therefore was not added to the model. cSUP had multiple standardized residual covariances greater than |2.58|,

indicating model misspecification (iREL3 = 3.48, ScScr = -5.40, NormRef = -5.80). While cSUP had a factor loading greater than .4 and was statistically significant, the factor and the corresponding error term was removed from the model (Model 4, χ^2 = 300.68, *p* = .000, df = 122, GFI =.933, CFI = .964, TLI = .955, RMSEA = .057), leaving five indicators of TSR (see Figure 26).

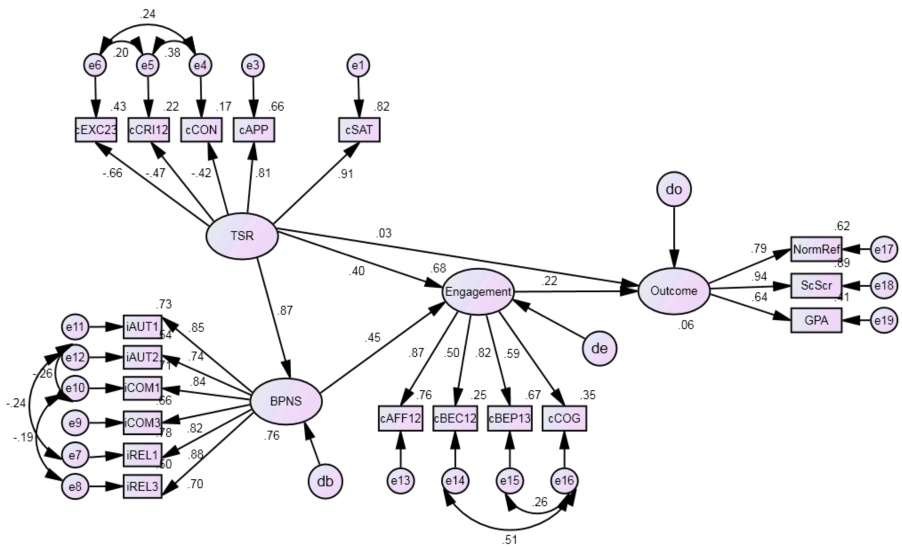


Figure 26. **Structural model regression weights and factor loadings**

Following the removal of cSUP, AMOS identified an improvement in the model if TSR and e19 (GPA) were set to covary. Modification indices of regression weights also showed significant improvement if all constructs and many indicators were allowed to load on GPA. GPA had twelve standardized residual covariances greater than |2.58|. In an exploratory fashion, TSR was allowed to covary with the error term for GPA. Modification indices regression weights no longer indicated better fit by allowing GPA to load on the other latent variables and indicator variables. Standardized residual covariance values were no longer an issue, and the model improved (χ^2 =254.509, *p* = .000, df = 122, GFI =.942, CFI = .973, TLI = .966, RMSEA = .049). Similarly, GPA was allowed to load on BPNS with identical model fit results and all standardized residual

covariances less than |2.58| (χ^2 =253.998, *p* = .000, df = 121, GFI =.942, CFI = .973, TLI = .966, RMSEA = .049).

The combined results of modification indices, standardized residual covariances, and the exploratory example described indicated a localized area of strain in the model, specifically with GPA. The simple remedy would be to remove GPA from the model; however, only two indicators of outcome would remain, which is not recommended. TSR and the error term for GPA could be set to covary or a path allowed to be freely estimated from BPNS to GPA, with the results ending in improved model fit according to fit indices and acceptable standardized residual covariances. The researcher, however, felt this was moving into exploratory SEM based on statistics and would result in an over fit model to the present data with no significant improvement in standardized regression weights (see Figure 27). GPA was left in the model.

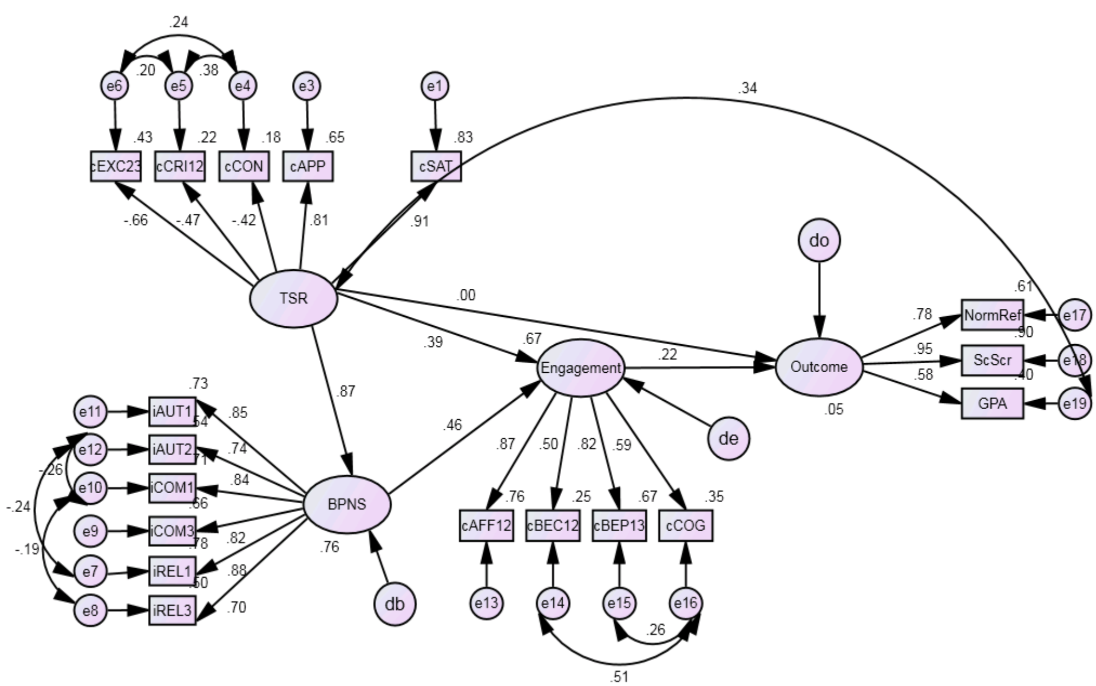


Figure 27. **Structural model regression weights and factor loadings with TSR and error term for GPA (e19) set to covary**

With GPA in the model, the only insignificant path was from TSR to outcome ($\beta = .03$, $p = .783$). While TSR had an indirect impact on outcome through engagement, Byrne (2014) recommended removing insignificant paths from the model and the path from TSR to outcome was removed (see Figure 28, Final SEM Model, $\chi^2 = 300.75$, $p = .000$, df = 123, GFI =.933, CFI = .964, TLI = .956, RMSEA = .056). Therefore, the final model was not statistically different from the model that included the path from TSR to outcome ($\Delta\chi^2 = -.07$, $\Delta df = 1$).

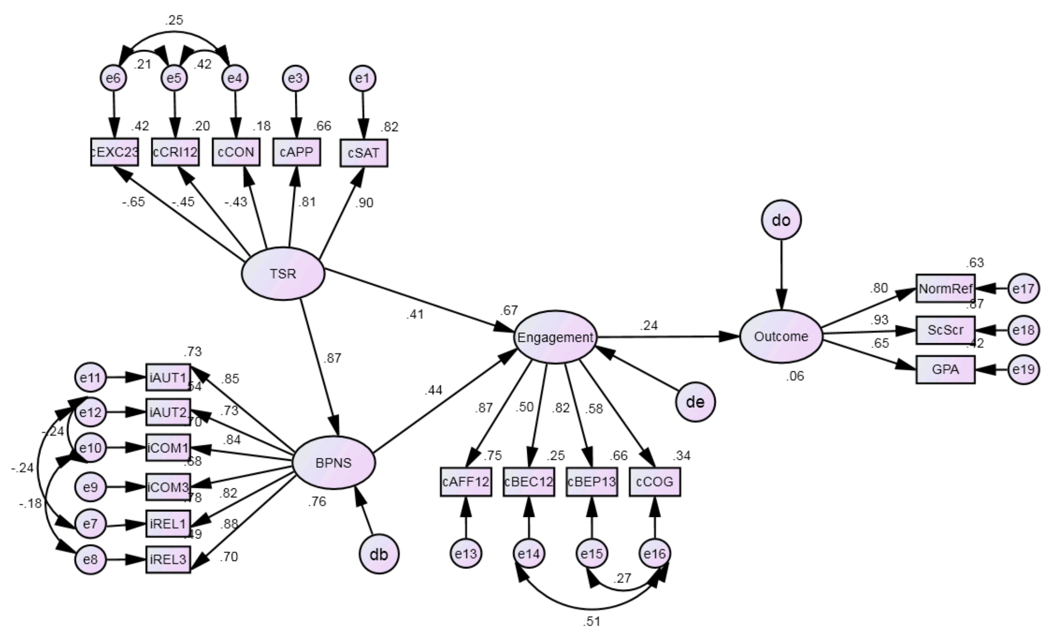


Figure 28. **Final structural model results with path from TSR to Outcome removed**

The final path model was not overly fit and satisfied the criteria of a model that had good global fit according to fit indices, reasonable regression weights in size and direction, acceptable sizes of standard errors, and acceptable standardized residual covariances with all variables except GPA. All modification indices that made

substantive sense were included in the model. All estimated regression weights, variances, and covariances between error terms were statistically significant. Findings were confirmed through Bayesian estimation, with near identical results (see Table 30 and 31).

Table 30

Full model unstandardized regression weight comparisons between Maximum Likelihood and Bayesian estimates

	ML Estimation Method				Bayesian Estimation Method		
	Est	S.E.	C.R.	P	Est	S.D.	C.R.
BPNS<---TSR	1.424	0.072	19.877	***	1.427	0.075	19.03

Table 30 Continued

	ML Estimation Method				Bayesian Estimation Method		
	Est	S.E.	C.R.	P	Est	S.D.	C.R.
Engagement<---TSR	0.423	0.102	4.165	***	0.424	0.103	4.12
Engagement<---BPNS	0.293	0.061	4.809	***	0.294	0.061	4.82
Outcome<---Engagement	9.556	2.021	4.728	***	9.567	2.067	4.63
iAUT1<---BPNS	1				1		
iAUT2<---BPNS	0.876	0.048	18.116	***	0.884	0.049	18.04
iCOM1<---BPNS	0.96	0.042	22.833	***	0.968	0.042	23.05
iCOM3<---BPNS	0.885	0.04	22.043	***	0.893	0.042	21.26
iREL1<---BPNS	1.021	0.041	25.107	***	1.028	0.043	23.91

iREL3<---BPNS	0.836	0.048	17.408	***	0.841	0.049	17.16
cAFF12<---Engagement	1				1		
cBEC12<---Engagement	0.345	0.032	10.62	***	0.348	0.033	10.55
cBEP13<---Engagement	0.815	0.041	19.644	***	0.815	0.043	18.95
cCOG<---Engagement	0.451	0.035	12.758	***	0.452	0.036	12.56
GPA<---Outcome	0.113	0.008	13.554	***	0.113	0.009	12.56
NormRef<---Outcome	0.448	0.028	16.187	***	0.449	0.028	16.04
ScScr<---Outcome	1			***	1		

Table 30 Continued

	ML Estimation Method				Bayesian Estimation Method		
	Est	S.E.	C.R.	P	Est	S.D.	C.R.
cAPP<---TSR	0.833	0.037	22.312	***	0.835	0.038	21.97
cCON<---TSR	-0.317	0.035	-9.04	***	-0.318	0.035	-9.09
cCRI<---TSR	-0.376	0.036	-10.383	***	-0.375	0.038	-9.87
cEXC23<---TSR	-0.635	0.04	-15.958	***	-0.636	0.04	-15.90
cSAT<---TSR	1				1		

Table 31

Full model standardized regression weight comparisons between Maximum-likelihood and Bayesian estimates

	ML Estimation Method	Bayesian Estimation Method	Difference in Estimation Method

BPNS<---TSR	0.872	0.873	-0.001
Engagement<---TSR	0.399	0.399	0
Engagement<---BPNS	0.451	0.447	0.004
Outcome<---Engagement	0.245	0.239	0.006
iAUT1<---BPNS	0.854	0.852	0.002
iAUT2<---BPNS	0.735	0.731	0.004
iCOM1<---BPNS	0.84	0.837	0.003
iCOM3<---BPNS	0.815	0.815	0

Table 31 Continued

	ML Estimation Method	Bayesian Estimation Method	Difference in Estimation Method
iREL1<---BPNS	0.883	0.882	0.001
iREL3<---BPNS	0.704	0.703	0.001
cAFF12<---Engagement	0.869	0.867	0.002
cBEC12<---Engagement	0.496	0.495	0.001
cBEP13<---Engagement	0.82	0.819	0.001
cCOG<---Engagement	0.588	0.587	0.001
GPA<---Outcome	0.639	0.639	0
NormRef<---Outcome	0.785	0.784	0.001

ScScr<---Outcome	0.943	0.943	0
cAPP<---TSR	0.811	0.809	0.002
cCON<---TSR	-0.417	-0.413	-0.004
cCRI12<---TSR	-0.47	-0.466	-0.004
cEXC23<---TSR	-0.656	-0.652	-0.004
cSAT<---TSR	0.908	0.904	0.004

Convergence statistic 1.0019 with 500 + 60,099 * 2 samples

While other structural models may show good fit with this dataset, the final model in this research had good model statistics and provided support for the Self-systems Process model. All factor loadings of observed indicators were statistically significant and greater than .4 and adequately reflected the underlying latent constructs. Standardized regression weights between latent constructs were also statistically significant and positive. Context (TSR) influenced self (BPNS), which included action (engagement), and consequently, influenced outcome (outcome).

To answer research question three, to what extent does the teacher-student relationship influence level of student engagement, the model in Figure 29 was used to identify standardized direct and indirect effects between TSR and Engagement. A more positive teacher-student relationship led to a higher level of engagement.

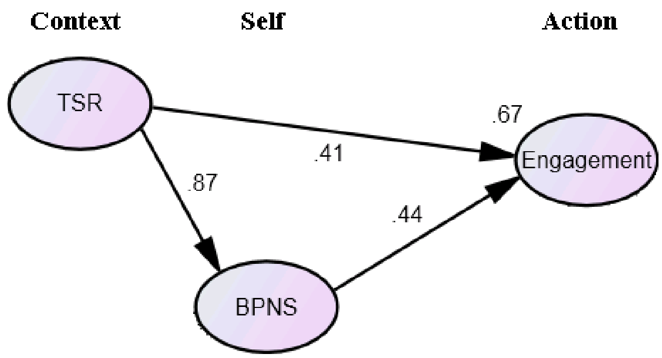


Figure 29. **Associations between Context, Self, and Action**

TSR and BPNS explained 67.5% of the variance in engagement with TSR explaining 76% of the variance in BPNS. The standardized direct effects of TSR on Engagement (β = .413) and BPNS (β = .872) were large, positive, and statistically significant similar to the effect of BPNS on engagement (β = .436). The effect of TSR on engagement was partially mediated by the latent variable BPNS. According to Kenny, Kashy, & Bolger, (1998), the amount of mediation is equal to the indirect effect when the mediator is included in the model. The indirect effect of TSR on engagement when mediated by BPNS was β = .380 (.872 x .436). The total standardized effect of TSR on Engagement (β = .783) was determined by adding both the direct and indirect effects, had a large impact on engagement, and indicated the importance of the teacher-student relationship psychological need satisfaction and student engagement..

There was no statistically significant difference in effect between LowSES and HighSES groups, and the indicators of BPNS and engagement were equal across groups. There were differences between the White and NonWhite groups with only cSAT, cAPP, and cCRI12 invariant across groups, which impacted the results of the structural model. A larger percentage of the variance in engagement was explained by TSR and BPNS for

the NonWhite (72.8%), as compared to White (66.7%) group, and the total effect of TSR on Engagement for the White group (.781) was lower than for the NonWhite group (.819), though not statistically significant.

Growth score structural model testing. The ScScr indicator was removed from the model and replaced with growth. The new measurement model of outcome consisted of the indicators NormRef, Growth, and GPA and could not be estimated as the model had zero degrees of freedom. Cronbach's alpha (.44) was much lower than when ScScr was included in the model (.69). The full structural model was run with adequate model fit (see Figure 30, X^2 = 303.734, p = .000, df = 123, GFI = .932, TLI = .950, CFI = .960, RMSEA = .057). All parameters were statistically significant with the exception of the factor loading from outcome to growth (β = .06, p = .306). There were three standardized residual covariances greater than |2.58| between growth and NormRef (2.682), growth and cEXC23 (-3.061), and GPA and iCOM3 (2.627). With scale score in the structural model, GPA was the area of misfit; however, with growth in the structural model, the growth indicator was the area of misfit. There were no substantive corrections based on the modification indices.

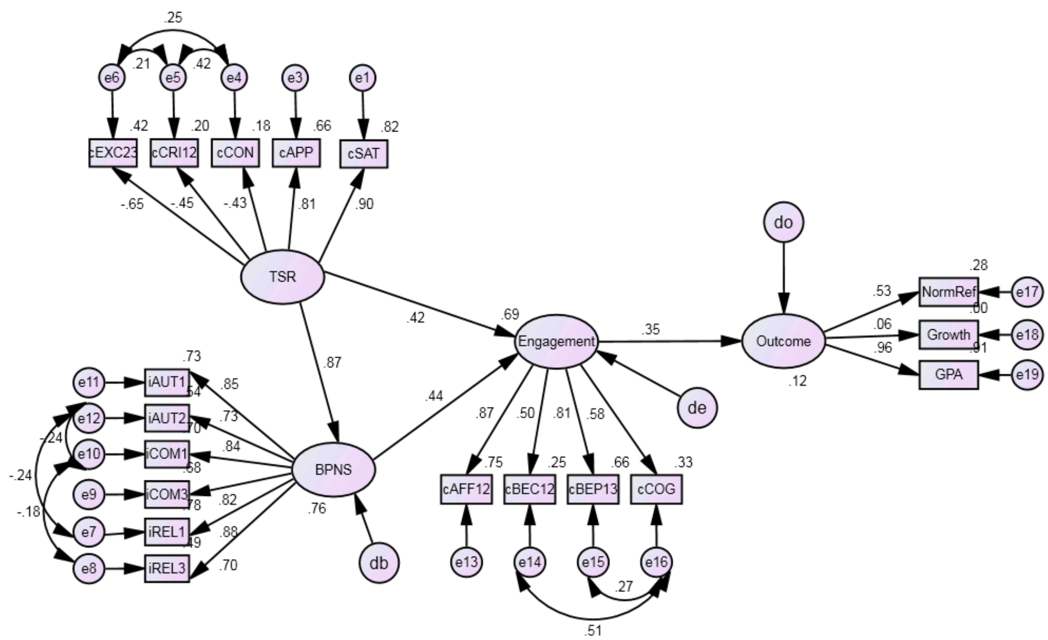


Figure 30. **Structural model results with growth as an indicator of outcome**

The purpose of research question one was to examine to what extent the teacher-student relationship influenced satisfaction of basic psychological needs which influenced engagement and consequently influenced student growth percentiles as compared to student status scores using an identical methodological setup (Context → Self → Action → Outcome). The two near identical structural models, one with growth as an indicator of outcome and one with scale score as an indicator of outcome, were compared. The structural model with growth had similar results to the structural model with ScScr. The scale score model ($\chi 2$ = 291.807, p = .000, df = 123, GFI = .934, TLI = .957, CFI = .966, RMSEA = .055) had a slightly better fit than the growth model ($\chi 2$ = 303.734, p = .000, df = 123, GFI = .932, TLI = .950, CFI = .960, RMSEA = .057). All paths and covariances were significant in both models, with the sole exception of the path from outcome to growth (p = .306). In the scale score model, GPA was an area of localized strain on the model, whereas in the model with growth, growth itself was the

localized area of strain. An examination of the overall models showed showed little difference in the constructs of TSR, BPNS, and engagement, as these were distal latent constructs (see Table 32).

Table 32

ScScr and growth standardized regression weight comparison of distal factors

Standardized Regression Weight		Standardized Regression Weight
Scale Score Model	Path	Growth Score Model
0.436	Engagement<---BPNS	0.443
0.853	iAUT1<---BPNS	0.854
0.732	iAUT2<---BPNS	0.734
0.839	iCOM1<---BPNS	0.838
0.825	iCOM3<---BPNS	0.825
0.881	iREL1<---BPNS	0.881
0.701	iREL3<---BPNS	0.702
0.872	BPNS<---TSR	0.872
0.81	cAPP<---TSR	0.808
-0.426	cCON<---TSR	-0.43
-0.452	cCRI12<---TSR	-0.453

Table 32

ScScr and growth standardized regression weight comparison of distal factors

Table 32 Continued

Standardized Regression Weight Scale Score Model	Path	Standardized Regression Weight Growth Score Model
-0.65	cEXC23<---TSR	-0.651
0.904	cSAT<---TSR	0.905
0.413	Engagement<---TSR	0.415
0.868	cAFF12<---Engagement	0.864
0.497	cBEC12<---Engagement	0.497
0.815	cBEP13<---Engagement	0.811
0.582	cCOG<---Engagement	0.578

As previously described, TSR had a large direct effect on BPNS and a large total effect on engagement when scale score was included in the model. Including BPNS, which had a significant impact on engagement, 67.5% of the variance in engagement was accounted for by TSR and BPN. When scale score was switched out with growth indicator, there was very little impact on the constructs of TSR, BPNS, and engagement and the findings as this part of the model was distal to the indicators of outcome. A slightly larger amount of variance was accounted for (68.9%) in engagement.

The significant difference between the two models was the regression weight from engagement to outcome and the indicators of outcome (see Table 33). Engagement explained less variance of outcome in the scale score model (6%) than it did in the growth model (11.9%) and had a lower standardized regression weight of β = .245 compared to β = .345. The association between engagement and outcome increased in the structural model that included the growth indicator; however, the increase was the result of the growth indicator being non-significant and more of the variance of outcome being accounted for by GPA.

Table 33

ScScr and growth standardized regression weight comparison of proximal factors

Standardized Regression Weight Scale Score Model	Path	Standardized Regression Weight Growth Score Model
0.245	Outcome<---Engagement	0.344
0.649	GPA<---Outcome	0.967
0.933	ScScr<---Outcome	
	Growth<---Outcome	0.054*
0.795	NormRef<---Outcome	0.518

*Non significant path

Whereas GPA was the area of misfit in the model that included ScScr, growth was the area of misfit in the model with growth. Outcome accounted for 87.1%, 42.1%,

and 63.3% of the variance in ScScr, GPA, and NormRef, respectively, in the scale score model and .4%, 91.4%, and 28.1% in growth, GPA, and NormRef, respectively, in the growth model. The factor loading of GPA increased in the growth model (.96), as compared to the ScScr model (.65) because growth was non-significant.

According to Byrne (2011), nonsignificant factors should be deleted from the model. The low factor loading of growth along with the low reliability coefficient of the construct of outcome indicated that growth was not a reliable indicator of outcome in this model. The growth indicator did not have a significant association with outcome and the other indicators of outcome; therefore, growth did not truly fit this model.

Prior research has shown that TSR, BPNS, and Engagement influence student outcomes. Results of this research confirm much of the prior research with the indicators of ScScr, GPA, and NormRef. When student growth percentiles were included in the model, a greater percentage of variance was accounted for by engagement; however, the growth indicator was not statistically significant. The totality of the evidence provided by the structural models indicated growth did not fit in this model and TSR, BPNS, and engagement did not influence growth. If the model was a valid measurement of the Self-systems Process model, then TSR, BPNS, and engagement had no significant impact on growth as determined by the State of Georgia for this dataset.

Multigroup Invariance - LowSES and HighSES groups

In SEM, it is possible to analyze multiple groups simultaneously to determine if the model is equivalent across groups (Byrne, 2004; Hox & Becher, 2004). Multigroup testing was utilized to address research question two: To what extent is the effect of teacher-student relationships on student growth percentiles invariant across low

socioeconomic status students and high socioeconomic status students, and to what extent is the effect of teacher-student relationships on student growth percentiles invariant across white students versus non-white students.

Prior to multigroup testing, descriptive statistics were examined to ensure assumptions of SEM were satisfied. The means, skewness, and kurtosis values were closely related in all variables and in the same direction with few exceptions (see Table 34). Interpretation of survey results showed LowSES students as having lower levels of competence, while also having much lower means for NormRef (ΔM = 15.457) and GPA (ΔM = 4.267). Skewness and kurtosis values were within |2| and |7|, respectively, for both groups. Univariate outliers were previously addressed. Two multivariate outliers were removed from the LowSES group and four from the HighSES group. Multivariate normality assessed with Mardia's coefficient indicated the data for the LowSES (6.738) group was more multivariate normal than the HighSES (14.070). Both univariate and multivariate statistics indicated maximum-likelihood estimation was an acceptable method for testing.

Table 34

HighSES and LowSES descriptive statistics

	All responses			HighSES			LowSES		
Indicator	Mean	Skew.	Kurt.	Mean	Skew.	Kurt.	Mean	Skew..	Kurt.
cAPP	3.209	-0.139	-0.893	3.191	-0.110	-0.842	3.249	-0.208	-1.015
cSAT	3.636	-0.544	-0.693	3.691	-0.563	-0.664	3.511	-0.477	-0.799
cCRI12	1.614	1.564	2.178	1.590	1.571	2.065	1.669	1.562	2.485

Table 34

HighSES and LowSES descriptive statistics

cEXC23	1.869	1.214	0.715	1.872	1.219	0.767	1.863	1.216	0.663
cCON	1.563	1.803	2.890	1.561	1.827	3.098	1.568	1.766	2.584
iAUT1	4.611	-0.352	-1.039	4.658	-0.414	-0.976	4.504	-0.221	-1.145

Table 34 Continued

	All responses			HighSES			LowSES		
Indicator	Mean	Skew.	Kurt.	Mean	Skew.	Kurt.	Mean	Skew..	Kurt.
iAUT2	4.165	-0.070	-1.214	4.225	-0.124	-1.167	4.029	0.053	-1.288
iCOM1	4.800	-0.497	-0.891	4.870	-0.598	-0.722	4.640	-0.284	-1.158
iCOM3	4.892	-0.541	-0.795	5.006	-0.672	-0.605	4.633	-0.267	-1.027
iREL1	4.785	-0.512	-0.948	4.854	-0.549	-0.874	4.626	-0.417	-1.111
iREL3	3.538	0.257	-1.168	3.566	0.206	-1.102	3.475	0.358	-1.297
cAFF12	3.536	-0.512	-0.952	3.484	-0.483	-0.947	3.655	-0.605	-0.922
cBEP13	3.807	-0.689	-0.519	3.809	-0.646	-0.590	3.802	-0.771	-0.411
cBEC12	4.467	-1.650	2.334	4.505	-1.705	2.422	4.381	-1.537	2.144
cCOG	4.165	-1.052	0.357	4.146	-1.027	0.279	4.207	-1.129	0.614
NormRef	69.908	-0.875	0.027	74.630	-0.939	0.415	59.173	-0.390	-0.977
Growth	56.024	-0.255	-1.152	55.519	-0.212	-1.162	57.173	-0.355	-1.118
GPA	87.165	-0.761	-0.069	88.468	-0.809	0.220	84.201	-0.435	-0.790

The final model from SEM testing with growth incorporated was used for multigroup testing using the automated process provided by AMOS. The model was used as the baseline model and was estimated simultaneously with both LowSES and HighSES groups, had no constraints, was used to compare all subsequent tests of invariance, and was termed the configural model. Configural invariance was tested to see the extent to which the structural model and indicators were similar across both groups. Low SES and high SES groups were set up in AMOS and the simultaneous estimation process was run with model fit indices showing decent model fit across both groups (see Table 35, Model 1, χ^2 = 508.295, p = .000, df = 246, CFI = .943, RMSEA = .049).

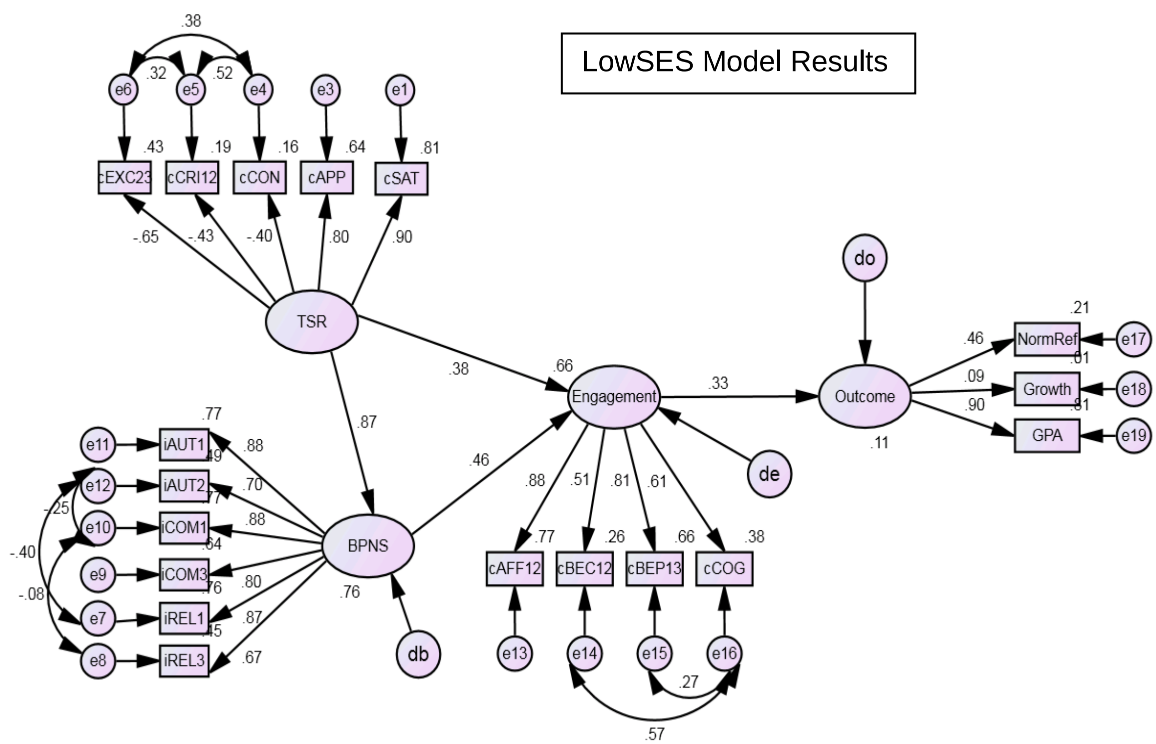
Table 35

Multigroup testing between HighSES and LowSES students model fit indices

Model	$\chi 2$	*df*	$\Delta\chi 2$	Δdf	Statistical significance	RMSEA	Notes
1	508.295	246	-	-	-		Configural Model
2	531.346	260	23.051	14	.059		Measurement weights constrained equal
3	531.935	264	23.640	18	.167		Measurement weights and structural weights constrained equal
4	533.109	265	24.814	19	.167		Measurement weights, structural weights, and structural covariances constrained equal

As a first step in testing equivalency, Byrne (2004) recommended to test the fully constrained model by constraining all factor loadings, factor variances, and factor covariances equal across the groups. Structural and measurement residuals were not

included in the analysis because, according to Byrne (2004), this was too restrictive a test. The chi-square difference between LowSES and HighSES groups was not statistically significant in any of the multigroup tests; however, it was nearly significant for the measurement weights test (*p* = .059). Multigroup testing revealed the full model to be invariant across LowSES and HighSES groups when constraining measurement weights, structural weights and structural covariances equal ($\Delta\chi2$ = 24.814, df= 19, *p* = .167) and no further invariance testing was needed (see Figure 31).



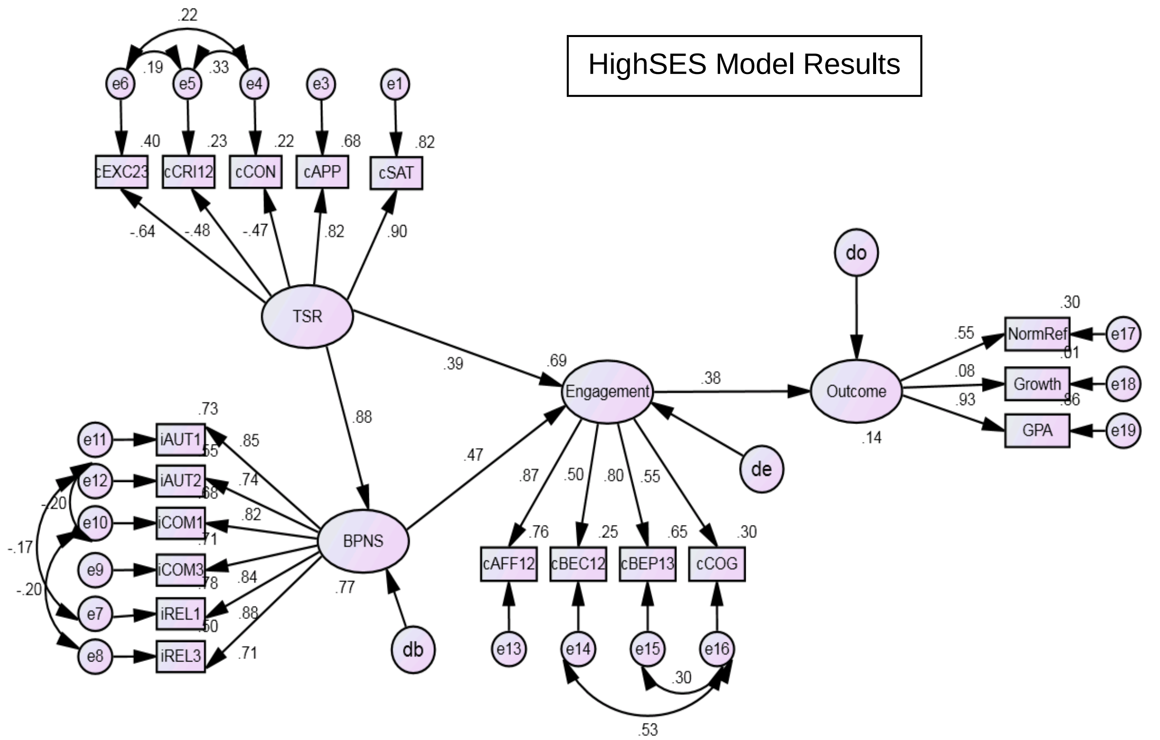


Figure 31. **Multigroup testing structural model results of HighSES and LowSES groups**

Multigroup Invariance - White and NonWhite groups

Descriptive statistics were examined for the White and NonWhite groups. The means, skewness, and kurtosis values were closely related in all variables and in the same direction with few exceptions (see Table 36). Nonwhite students responded as having lower levels of discord with their teachers than their counterparts and had a less peaked distribution for behavioral engagement compliance. NonWhite students had lower means for NormRef (ΔM = -6.458) and GPA (ΔM = -1.981) and slightly higher level of growth (ΔM = 1.981). Skewness and kurtosis values were within acceptable range; however, cCRI12 (*sk* = 2.148, *k* = 6.120) was borderline. Univariate outliers were previously addressed and neither group had multivariate outliers. Multivariate normality assessed with Mardia's coefficient indicated the data for the NonWhite (10.564) group was

slightly more multivariate normal than the White (11.202) group, with both groups being near normal. Both univariate and multivariate statistics indicated maximum-likelihood estimation was an acceptable method for testing.

Table 36

White and NonWhite descriptive statistics

	All responses			White			NonWhite		
Indicator	Mean	Skew.	Kurt.	Mean	Skew.	Kurt.	Mean	Skew.	Kurt
cAPP	3.209	-0.139	-0.893	3.154	-0.145	-0.862	3.352	-0.163	-0.991
cSAT	3.636	-0.544	-0.693	3.643	-0.598	-0.537	3.619	-0.427	-1.016
cCRI12	1.614	1.564	2.178	1.688	1.376	1.393	1.420	2.148	6.120
cEXC23	1.869	1.214	0.715	1.895	1.190	0.717	1.800	1.308	0.840
cCON	1.563	1.803	2.890	1.617	1.703	2.390	1.421	1.924	3.537

Table 36 Continued

	All responses			White			NonWhite		
Indicator	Mean	Skew.	Kurt.	Mean	Skew.	Kurt.	Mean	Skew.	Kurt
iAUT1	4.611	-0.352	-1.039	4.515	-0.327	-1.054	4.864	-0.435	-0.988
iAUT2	4.165	-0.070	-1.214	4.173	-0.033	-1.167	4.144	-0.145	-1.339
iCOM1	4.800	-0.497	-0.891	4.779	-0.503	-0.835	4.856	-0.496	-1.012
iCOM3	4.892	-0.541	-0.795	4.848	-0.557	-0.769	5.008	-0.479	-0.935
iREL1	4.785	-0.512	-0.948	4.776	-0.532	-0.911	4.808	-0.466	-1.033
iREL3	3.538	0.257	-1.168	3.464	0.273	-1.114	3.736	0.191	-1.319

cAFF12	3.536	-0.512	-0.952	3.480	-0.480	-0.991	3.684	-0.590	-0.861
cBEP13	3.807	-0.689	-0.519	3.767	-0.596	-0.647	3.912	-0.957	-0.025
cBEC12	4.467	-1.650	2.334	4.459	-1.682	2.621	4.488	-1.574	1.588
cCOG	4.165	-1.052	0.357	4.120	-1.020	0.282	4.282	-1.144	0.578
NormRef	69.908	-0.875	0.027	71.682	-0.934	0.212	65.224	-0.723	-0.349
Growth	56.024	-0.255	-1.152	55.312	-0.193	-1.257	57.904	-0.426	-0.824
GPA	87.165	-0.761	-0.069	87.709	-0.769	0.126	85.728	-0.656	-0.592

Configural invariance was tested to see the extent to which the structural model and indicators were similar across both groups. White and NonWhite groups were set up in AMOS and the simultaneous estimation process was run with model fit indices showing decent model fit across both groups (see Table 37, Configural Model, χ^2 = 446.014, p = .000, df = 246, CFI = .956, RMSEA = .043). The model was then fully constrained to include measurement weights, structural weights, and factor variances and was not invariant across White and NonWhite groups (Model 1, $\Delta\chi 2$ = 32.53, Δdf = 19, p = .027).

To assess metric invariance, factor loadings were constrained equal for TSR (Model 2, $\Delta\chi 2$ = 19.704, Δdf = 4, p = .001), BPNS (Model 3, $\Delta\chi 2$ = 5.100, Δdf = 5, p = .404*), engagement (Model 4, $\Delta\chi 2$ = 5.264, Δdf = 3, p = .153*), and outcome (Model 5, $\Delta\chi 2$ = 1.731, Δdf = 2, p = .421*) individually, with all other measurement models estimated freely to determine if groups were invariant (Templin, 2012) across instruments. All measurement instruments were invariant across groups (*) except TSR.

Individual indicators of TSR were then constrained equal one at a time to identify which indicators were invariant (see Table 37). cSAT was already constrained at 1 and was invariant across groups. cAPP (Model 2a) and cEXC (Model 2d) were also determined to be invariant across groups. All measurement weights except cCON and cCRI12 were constrained equal with invariant findings (Model 6, $\Delta\chi 2$ = 12.194, Δdf = 12, *p* = .430*). Structural weights were then constrained equal (Model 6a, $\Delta\chi 2$ = 13.094, Δdf = 16, *p* = .666*) followed by structural covariances (Model 6b, $\Delta\chi 2$ = 13.501, Δdf = 17, *p* = .702*).

Table 37

Multigroup testing between White and NonWhite students model fit indices

Model	χ 2	*df*	$\Delta\chi$ 2	Δ*df*	Statistical significance	RMSEA	Model Description
	446.014	246	-	-	-	.956	Configural Model - Baseline
1	478.547	265	32.532	19	*p* = .027*	.953	Measurement weights, structural weights, and structural covariances constrained equal

Table 37 Continued

Model	χ 2	*df*	$\Delta\chi$ 2	Δ*df*	Statistical significance	RMSEA	Model Description
2	465.718	250	19.704	4	*p* = .001	.952	All TSR measurement weights constrained equal
2a	446.107	247	.093	1	*p* = .761*	.956	cAPP constrained equal, invariant
2b	454.676	248	8.661	2	*p* = .013	.954	cCON constrained equal, not invariant
2c	460.216	248	14.202	2	*p* = .001	.953	cCRI12 constrained equal, not invariant
2d	446.190	248	.175	2	*p* = .916*	.956	cEXC23 constrained equal, invariant

3	451.114	251	5.100	5	*p* = .404*	.955	BPNS measurement weights constrained equal. Invariant across groups
3a	451.285	253	5.270	7	*p* = .627*	.956	cAPP, cEXC23 and all weights of BPNS constrained equal. Invariant across groups
4	451.278	249	5.264	3	*p* = .153*	.955	ENG measurement weights constrained equal. Invariant across groups
4a	456.545	256	10.531	10	*p* = .395*	.955	cAPP, cEXC23, BPNS, and all weights of engagement constrained equal. Invariant across groups
5	447.745	248	1.731	2	*p* = .421*	.956	Outcome measurement weights constrained equal. Invariant across groups
5a	458.208	258	12.194	12	*p* = .430*	.955	cAPP, cEXC23, BPNS, engagement, and all weights of outcome constrained equal. Invariant across groups.
6	458.208	258	12.194	12	*p* = .430*	.955	Measurement weights without cCRI12 and cCON
6a	459.109	262	13.094	16	*p* =.666*	.956	Measurement weights and structural weights without cCRI and cCON included

Table 37 Continued

Model	χ 2	*df*	Δ χ 2	Δ*df*	Statistical significance	RMSEA	Model Description
6b	459.516	263	13.501	17	*p* =.702*	.956	Measurement weights, structural weights, and structural covariances without cCRI12 and cCON included

* Not Statistically significant. Groups equal

White and NonWhite groups were invariant across the full structural model with the exception of cCON and cCRI (see Figures 32), in which the NonWhite group indicated statistically significant lower levels of conflict (β = -.29) and criticism (β = -.33) as compared to the White group (cCon β = -.47, cCRI β =-.48).

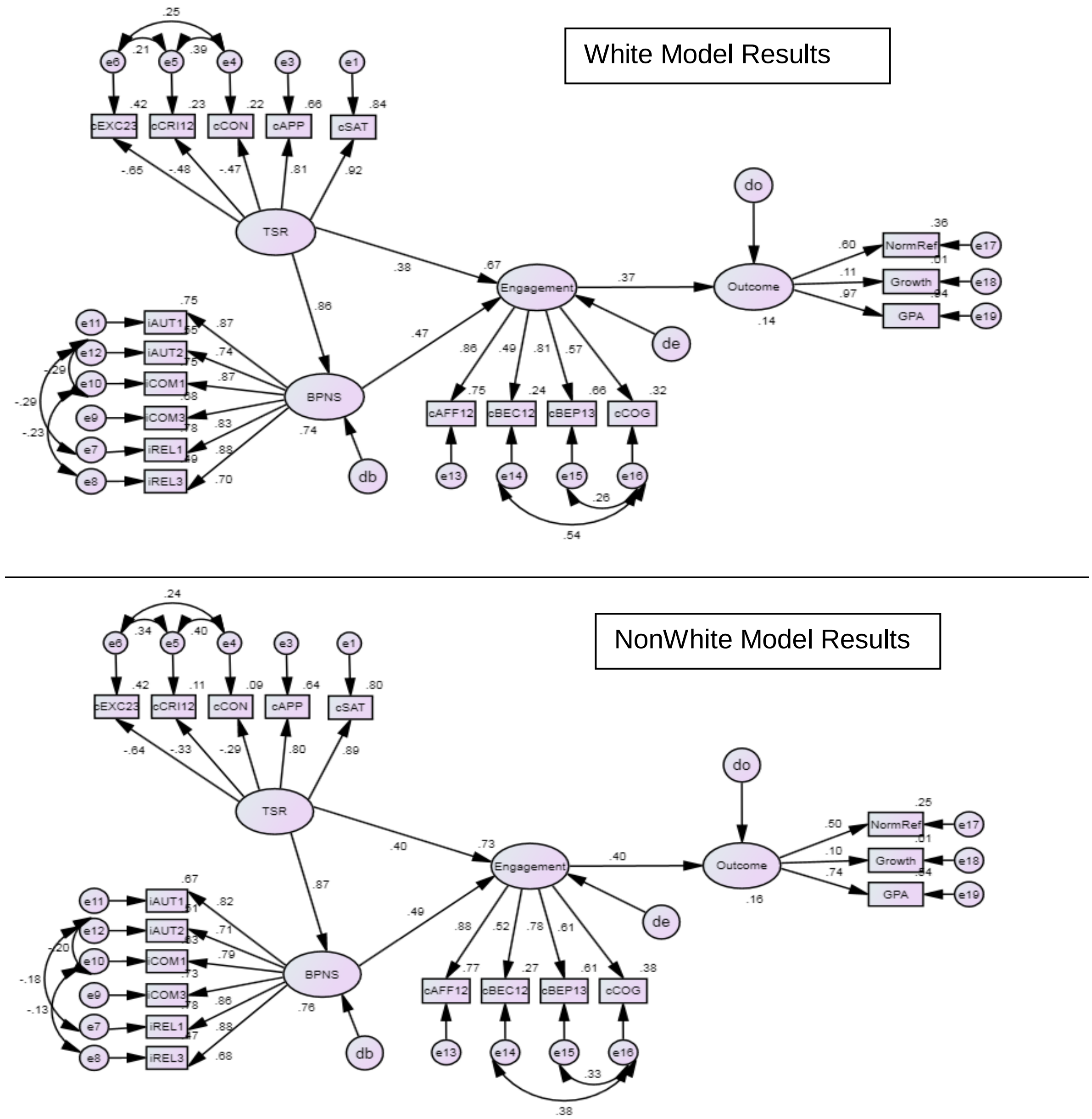


Figure 32. **Multigroup testing structural model results of White and NonWhite groups**

The purpose of research question two was to examine the extent the effect of teacher-student relationships on student growth percentiles was invariant across population subgroups (i.e. Low socioeconomic status students versus high socioeconomic status students and White students versus non-white students). TSR, BPNS, and engagement did not influence growth in the hypothesized structural equation model.

Invariance testing revealed similar results in that, while there was good global model fit, the path from outcome to growth was not significant (p = .179) across the LowSES and HighSES groups. A low amount of variance in growth was explained by the latent variable outcome for both the LowSES (.9%) and HighSES (.7%) groups. The path from outcome to growth was not significant (p = .055) across the White and NonWhite groups when constraining all measurement weights except cCON and cCRI12, structural weights, and structural covariances equal across groups. A low amount of variance in growth was explained by the latent variable outcome for both the White (1.3%) and NonWhite (1.0%) groups. There was no difference in the effect on growth.

Summary

Using a modified version of the Network of Relationships inventory, Basic Psychological Needs inventory, and the Classroom Engagement inventory, a total of 543 response sets were collected at a rural middle school in southwest Georgia using Google forms. The datasets were screened for abnormalities and missing data with 4.3% of the datasets removed due to missing scale or growth score values due to students not completing the Georgia Milestones assessment. Univariate and multivariate outliers were removed from the dataset.

In general, students responded to the measurement instruments as having a positive teacher-student relationship, having their basic psychological needs met, and being actively engaged in the classroom. Descriptive statistics revealed some inventory items were not necessarily appropriate for this research and reverse worded questions had atypical results. While the histograms of many of the indicator items did not appear

normal, all items had skewness and kurtosis values less than |2| and |7| respectively which made maximum-likelihood an appropriate estimation method.

During testing, AMOS presented non-positive definite matrices and negative variances which were inadmissible solutions. Constructs and indicators were examined for multicollinearity using correlations, variance inflation factor scores, and tolerance scores which resulted in the constructs of closeness and discord being collapsed into TSR and the constructs of autonomy, competence, and relatedness being collapsed into BPNS. Term4Avg was also dropped from the model as an indicator of outcome as the correlation, variance inflation factor score, and tolerance statistic indicated Term4Avg was too closely related to GPA.

Each inventory was validated through CFA initially with unparceled indicators. Descriptive statistics and information obtained during CFA were used to remove items that presented issues. The remaining items of the NRI and CEI were used to create parcels, that were used in measurement models, which were also validation through CFA. The measurement models of BPNS and Outcome were also analyzed, with all construct measurement models having adequate model fit. The measurement models were then input in the full structural model for model estimation and testing.

The full structural model was estimated with modification guided by model fit indices, modification indices of covariances and regression weights, and inspection of standardized residual covariances. The final structural model had good global fit; however, had a localized area of strain with GPA. The only path that was not statistically significant ran from TSR to outcome and was removed. The change had no statistically

significant impact on the structural model. All findings of the measurement models and structural model were verified and had similar findings using Bayesian estimation.

Structural equation modeling is a process used to confirm the plausibility of a model (Byrne, 2010; Kline, 2011). The findings of this research demonstrated the model, based on the SSPM, were plausible, and had good fit indices, regression weights of proper size and direction, and minimal residual issues. As SEM is a confirmatory process, it was possible that other competing models could be just as plausible with this dataset. The findings showed engagement was highly influenced by both TSR and BPNS. TSR had both direct and indirect effects on engagement as TSR was mediated by BPNS. TSR also had a large direct effect on BPNS. The impact of engagement on outcome, while not large, was statistically significant. The indicators of outcome had a large portion of their variance accounted for by outcome.

Scale score was removed from the model and replaced by growth for comparison purposes. Results of the slightly modified model were similar to the original model in the distal section, which included TSR, BPNS, and engagement. The major difference between the models was on the outcome construct. In the original model, GPA was the strain on the model, but now student growth percentiles were the strain. While the path from engagement to outcome was still significant, the growth indicator was not. This model, while still having acceptable fit indices, did not fit the dataset as well when scale score was included. According to the model, TSR, BPNS, and engagement did not influence growth as compared to scale score.

Multigroup testing revealed the structural model with growth as an indicator was invariant across the LowSES and HighSES and invariant across the White and NonWhite

groups when constraining all structural covariances, structural weights, and measurement weights except cCON and cCRI12 equal. Again, there was no impact of TSR, BPNS, or engagement on growth.

CHAPTER V
DISCUSSION

Summary

Recent changes in educational law have changed how districts, schools, administrators, and teachers are held accountable. Starting with the standards-based accountability system NCLB, status scores on state assessments were the sole criteria by which districts were rated. Students had to reach an increasing level of proficiency year after year on standardized assessments or were considered failing and counted against the district or school. It was possible that a student did not have the prerequisite skills needed to be successful on the assessments, yet was required to take the assessment,

which many times led to poor results. The focus of these accountability systems was on the district and individual schools. There was no teacher accountability.

The next round of educational reform, Race to the Top (RttT), shifted the focus of accountability from districts and schools, to school leaders and teachers, based on the performance of their students. RttT included the requirement that school leader and teacher evaluation systems incorporate growth as part of the overall evaluation system. In the State of Georgia, the new multidimensional evaluation system, Teacher Keys Effectiveness Systems (TKES) consisted of administrator evaluations, student surveys, and student growth determined using student growth percentiles.

Student growth percentiles are not growth based on gain scores, but based on how a student progressed compared to his or her academic peers on prior year's test compared to the current year's test. A student's growth does not just depend on how well he or she does, but how well other comparable students throughout the State do. A student may have exceeded proficiency levels on an assessment, yet have low growth as the other comparable students may have far exceeded proficiency levels.

Prior to implementation of TKES, students scores, status or growth, played no part in an educator's overall evaluation. The new growth metric was now a major component of a teacher's evaluation and accounted for 50% of the teacher rating. While there is a plethora of evidence linking educator practices to student achievement based on status scores, there is little peer reviewed research on classroom variables and how they impact student growth as it pertains to student growth percentiles.

Research has shown that positive teacher-student relationships and satisfying a student’s basic psychological needs influence a student's level of engagement and

consequently, student outcomes such as class averages, GPA, teacher test scores, and standardized assessment results. With a portion of teacher and leader accountability now based on student growth as determined by student growth percentiles, research of factors, both proximal and distal, that may impact student growth is needed, as research of factors that influence student achievement may not be applicable.

The model used in this research was based on the Self-System Process Model (Connell & Wellborn, 1991), which is grounded in Self-Determination Theory, which has been shown to influence student achievement (Deci & Ryan, 1985). The linear SSPM identified that social context and environment (context) effect basic psychological needs, (self) which in turn influences a student's level of engagement (action) and, consequently, achievement (outcome) (Reschly & Christenson, 2012; Skinner et al., 2008; Skinner & Pitzer, 2012).

This research was driven by the lack of information available on the connection between classroom variables and student growth percentiles. The goal of this research was to determine the extent that teacher-student relationships and satisfaction of basic psychological needs influence engagement and achievement as measured with student growth percentiles. This research was based on and built off prior findings in the research linking teacher-student relationships, basic psychological needs satisfaction, and engagement on improved student outcomes. Structural equation modeling was the statistical tool utilized to examine relationships between constructs. The model included GPA, norm-referenced status scores, and scale status scores set as the dependent variable with the results then compared to the results of an identical methodological setup with student growth percentiles switched out with scale status scores.

Summary of Research Findings

There was support for the self-systems process model in that context (TSR) influenced self (BPNS) which influenced action (engagement) and, consequently, outcome (outcome) with either scale score or growth in the structural model. The hypothesized model was setup to be recursive with arrows indicating causation in one direction from context to outcome. TSR was hypothesized to affect BPNS; however BPNS was not hypothesized to affect TSR. The structural model was not overly fit, had good global fit indices, reasonable regression weights in size and direction, acceptable sizes of standard errors, acceptable standardized residual covariances with all variables except GPA, and statistically significant regression weights, variances, and covariances. While SEM is a confirmatory process in which a model is determined to be plausible, it is possible that other hypothesized models would fit the dataset used in this research.

TSR had a statistically significant effect on BPNS and engagement. 67.5% of the variance in engagement was accounted for by TSR and BPNS with 76% of the variance in BNPS accounted for by TSR. BPNS mediated the effect of TSR on engagement, with TSR having a standardized indirect effect of $\beta = .380$ and a total effect of $\beta = .783$. In the structural model, the effect of TSR on outcome was not statistically significant and was removed; however, engagement mediated the effect of TSR on outcome which had an indirect effect of $\beta = .194$.

When the scale score indicator was replaced with student growth percentiles, there was little to no impact on the constructs of TSR, BPNS, and engagement as they were distal to the indicators of outcome. TSR continued to have a statistically significant effect on BPNS and engagement with a slightly higher of the variance in engagement accounted for by TSR and BPNS (68.9%) and 76% of the variance in BNPS accounted

for by TSR. The total standardized effect of TSR on engagement increased to $\beta = .801$ as the direct effect increased ($\beta = .417$). BPNS continued to mediate the effect of TSR on engagement, with TSR having a standardized indirect effect of $\beta = .384$.

In the structural model with scale score included, the effect of TSR on outcome was not statistically significant and was removed; however, engagement mediated the effect of TSR on outcome which had an indirect effect of $\beta = .194$ as compared to when growth was in the model ($\beta = .277$). While a direct path from BPNS was not included in the hypothesized model, BPNS had an indirect effect on outcome ($\beta = .107$) with similar findings in the structural model with growth ($\beta = .152$).

Mutligroup testing revealed that the structural model with growth included was invariant across the LowSES and HighSES groups and was invariant across the White and NonWhite groups when constraining all structural covariances, structural weights, and measurement weights except cCON and cCRI12 equal. The classroom engagement instrument and the basic psychological needs instruments were invariant across all groups while the NRI was not.

The structural models with scale score and growth were compared to examine how teacher-student relationships influence student growth percentiles. As previously stated, the scale score model was plausible. The growth model too was plausible with good global fit indices, reasonable regression weights in size and direction, acceptable sizes of standard errors, and acceptable standardized residual covariances with all variables except growth. The two models were very similar in findings except when comparing the constructs of outcome. In the growth model, growth was not statistically significant ($p = .306$) and was a localized area of strain in the structural model whereas

scale score was statistically significant and had a large standardized regression weight (β = .933). A standardized unit increase in engagement was associated with a β = .24 standardized unit increase in outcome which accounted for 87.1% of the variance of scale score. In the growth model, a standardized unit increase in engagement was associated with a β = .36 standardized unit increase in outcome, which accounted for .4% of the variance of growth. So while the effect of engagement on outcome was larger, it was not due to student growth, which was not statistically significant. TSR, BPNS, and engagement had significant impact on student outcomes across groups, but did not impact student growth percentiles in this research.

Discussion of Research Findings

The results of the full structural equation analysis provided support for the full Self-system Process Model (SSPM) hypothesized by Connell and Wellborn (1991). Context (TSR) was positively associated with self (BPNS), which was positively associated with action (engagement), which was consequently associated with outcome (outcome). While the original SSPM was linear in nature with one factor acting directly on the factor next to it, evidence was provided in the research that context influenced self and action both directly and indirectly.

The investigation of TSR, BPNS, and engagement led to many of the same conclusions supported in previous research. Connell and Wellborn (1991) and Reschly and Christenson (2012) found there was a direct relationship between BPNS and engagement. According to the researcher's hypothesized model, there was a direct positive relationship between BPNS and engagement (β = .436). The results also confirmed the findings of Stroet et al., (2013) that student perception of psychological

needs influenced level of student engagement. While a direct path was not included from BPNS on outcome, BPNS had an indirect effect ($\beta = .107$) on outcome and was mediated by engagement for which Connell and Wellborn and Reschly and Christenson found evidence.

Connell and Wellborn (1991), using path analysis, found a direct relationship between engagement and achievement test scores. The model provided evidence of a relationship between engagement and outcome, specifically with student scale scores on the Georgia milestones assessment. Following the path from engagement to scale score, a 1 standardized unit increase in engagement was associated with a .24 standardized unit increase in outcome, which accounted for 87.1% of the variance in scale score. The model did not provide evidence of a relationship between engagement and growth. While the association between engagement and outcome was larger ($\beta = .360$) when growth was in the model, outcome accounted for only .4% of the variance in growth and was not statistically significant.

The hypothesized model was not set up to study bidirectional feedback loops between TSR and engagement, like the research of Reeve (2012) who found evidence to support feedback loops. This research was conducted towards the end of the school year using indicators of outcome from the end of the year state assessment and classroom grades. While no assumption is perfect, it could be assumed the end results of bidirectional feedback loops were captured in the research. The relationship between teachers and students had been forming and adjusting throughout the school year, impacting engagement, while a student's level of engagement has influenced the relationship between students and teachers. Results would have been different had the

questionnaire been completed a month after school started or around midyear. Collecting student responses near the end of school year captured the result of the bidirectional feedback loops; however, the research provided no evidence of bidirectional feedback loops.

Results of this research were similar to the findings of Furrer and Skinner (2003), who, in a sample of 641 third to sixth grade students, found support for the SSPM. Students who felt connected and supported and had a greater level of relatedness had higher levels of engagement, worked harder, and had more positive affect and greater academic success. Students that reported a more positive TSR reported higher levels of psychological need satisfaction and higher levels of engagement, which led to higher outcomes. Furrer and Skinner; however, did not include the BPNS in their model. Also, like the sample used in the Furrer and Skinner research (95% white), findings of this research may not be generalizable, as the sample was not very diverse, had a low percentage of low socioeconomic status students, and did not include elementary and high school students.

Roorda et al., (2011) in a meta-analytic review on TSR, engagement, and achievement, found large associations between TSR and engagement and a smaller association between TSR and achievement, which the results of this research also showed. The direct and indirect effects of TSR on engagement were both positive and large ($\beta = .783$). This research also found a small indirect effect of TSR on outcome ($\beta = .194$) in which the direct effect removed from the structural model as the path was not statistically significant. The effect of TSR on outcome was mediated by engagement similar to the findings of Roorda et al. When similar informants were used to report

levels of TSR and engagement, associations between the constructs were elevated, possibly due to shared variance of using the same informant (Furrer and Skinner 2003; Reyes et al., 2012; Roorda et al.). The same informants were utilized in this research; therefore, the total standardized effects between TSR and BPNS, BPNS and engagement, and TSR and engagement may be elevated.

Hamre and Pianta (2001) identified three dimensions to the TSR, which included closeness, dependency, and conflict and were found to be invariant across age, ethnicity, and socioeconomic status. While not identical to the indicators Hamre and Pianta used, the indicators of closeness, satisfaction and support, were found to be invariant across LowSES and HighSES groups and White and NonWhite groups when growth was included as an indicator of outcome. Exclusion, conflict, and criticism were invariant across LowSES and HighSES groups; however, only exclusion was invariant across White and NonWhite groups. The NonWhite group reported statistically significant lower levels of criticism (M = 1.42) and conflict (M = 1.421) as compared to the White group (M_{cCRI12} = 1.688 and M_{cCON} = 1.617) While the impact of the indicators of TSR on outcome were not directly studied, the factors of closeness, satisfaction (β = .9) and support (β = .8), had large factor loadings indicating their importance in the construct of TSR while the factors of discord, conflict (β = -.43), criticism (β = -.45), and exclusion (β = -.65) had much lower loadings. A greater amount of variance was accounted for in TSR through the closeness factors than the discord factors.

While levels of engagement generally decrease as students get older, Bingham and Okagaki (2012) noted ethnicity and socioeconomic status do not have such a simple relationship. There were many factors pertaining to self-identity, culture, family support,

teacher support, school makeup, and teacher race when trying to generalize engagement levels by race and socioeconomic status. Whereas Marks, (2000) and Wang et al., (2014) found that Low SES students consistently showed lower levels of engagement than their counterparts, measurement of engagement levels across socioeconomic status and race using the classroom engagement instrument were found to be invariant. Low SES students reported higher levels of affective and cognitive engagement, lower levels of behavioral engagement compliance, and the same level of behavioral engagement participation compared to High SES students. Both Wang et al., and Marks utilized samples of students from metropolitan areas with much larger school sizes. It is possible that the smaller more affluent school district used in this research impacted Low SES students' reports of engagement in the classroom.

Similar to the findings of Conner and Pope (2013), there were no differences in student reported levels of engagement between the White and NonWhite groups at the middle school level. Conner and Pope surveyed students from middle school to high school on their levels of engagement and found that behavioral engagement was self-reported highest by students, followed by cognitive and emotional engagement respectively. This research had the same findings with behavioral engagement compliance having the largest student reported means. The similar findings may have been a result of both samples being from high performing schools and school districts.

In the literature review, no peer reviewed research was identified, and only three dissertations were identified that investigated factors that influence student growth percentiles. Cervoni (2014) investigated many factors endorsed by New York State that have shown to improve student achievement, such as differentiated instruction, group

work, encouraging student engagement, use of formative assessments, and years of teaching experience with none of the factors influencing student growth percentiles. Unlike the current study which utilized student reported levels of TSR, BPNS, and levels of engagement, Cervoni utilized teacher reported levels of the indicators used in the study. Use of standards-based report cards has also been shown to improve levels of student achievement. Craig (2011) however, found use of standards based report cards had no impact on student growth percentiles. LeGeros (2013) focused on the relationship between student growth percentiles and elementary math teacher licensure exams in the state of Massachusetts. Students with teachers who conditionally and fully passed the MTEL had statistically significantly higher student growth percentiles than students with teachers who failed the MTEL test. Passing the MTEL state licensure exam showed a teacher had detailed content knowledge, and resulting instruction influenced student growth in the classroom.

Study Limitations

There were several limitations to this study. First, the study was very narrow in the population of choice as this was the first time TSRs were tested to determine if they influenced or were related to student growth percentiles. The sample was one of convenience from a middle school which served seventh and eighth grade students in a rural school district located in southwest Georgia with a relatively white affluent population. With such a limited scope, study findings may not be generalizable to students in grades K-6 or 9-12 with differing demographics.

Second, students were surveyed as to their perception of their relationships with their teachers, satisfaction of basic psychological needs, and their level of engagement in

the classroom. Because the survey was completed by students at a single point in time, it is possible that their perception that day was influenced by how they felt that day, good or bad. Student responses could have been influenced by any positive or negative interaction with their teacher the day the survey was administered.

Third, while the instruments were originally designed for students in grades four and higher, confirmatory factor analysis showed that students struggled to answer questions that were reverse coded. Students may have misunderstood other questions on the inventories, impacting the results of student responses.

Fourth, the construct of outcome was not well defined when student growth percentiles were included in the structural model. In the structural model that included scale score, outcome had the minimum number of recommended indicators, three, with good factor loadings and an acceptable reliability coefficient; however, in the model that included growth, only two of the three indicators had good factor loadings with growth having almost no influence on outcome and a low reliability coefficient (Cronbach's alpha = .341). The researcher believed that student growth percentiles were similar to other indicators of outcome like class average, GPA, teacher generated test scores, and standardized assessment scores. The results indicated that growth was not an indicator of outcome as defined in the literature. Stated differently, growth is not like traditional indicators of outcome such as exam scores, standardized test scores, or student GPAs.

Finally, the structural model with good fit is not the absolute model for the relationship between constructs under study. While the structural paths in the model are supported by prior research and the model had good fit indices, alternative models may do just a good of a job fitting the constructs under study.

Implications

Many of the research findings supported prior research on the relationships between teacher-student relationships, basic psychological need satisfaction, engagement, and student outcomes. In this study with this group of students using the NRI, BPNS, and CEI inventories, evidence was provided that TSR and BPNS were positively associated with engagement and, consequently, outcome. TSR had both a direct and indirect effect on engagement with the indirect effect working through BPNS. Engagement also mediated the effect of both TSR and BPNS on outcome as both indicators had small indirect effects on outcome. Findings were consistent across low socioeconomic status groups and high socioeconomic status groups and across white and non-white groups. The findings highlight that if educators want their students to be highly engaged in the classroom, they need to create a context that will promote a positive teacher-student relationship that satisfies a student's basic psychological needs.

Student growth now plays a tremendous role in teacher evaluations. In the era of new teacher accountability systems that incorporate student growth percentiles as part of the evaluation system, it is essential to recognize and understand factors that influence student growth to help both teachers and students excel. In comparing the structural models that included ScScr and Growth, it was determined that the factors of TSR, BPNS, and engagement did not positively or negatively affect student growth percentiles. Student growth had low correlations to all factors and constructs and there was little to no association between student growth percentiles and traditional indicators of outcome such as GPA and standardized test scores. Classroom practices that are known to improve student outcomes had no impact on student growth percentiles which raises the question of how teachers can improve their students growth.

Additionally, the research determined that growth did not fit in the construct of outcome using traditional indicators of outcome such as GPA and standardized test results. Student growth percentiles should not be used as indicators of outcome like student test scores or GPA. Student growth had low correlations to all factors and constructs, and there was little to no association between student growth percentiles and traditional indicators of outcome such as GPA and standardized test scores. If structural equation modeling is used to analyze factors that can influence growth percentiles in the future, it will be necessary to find other indicators that are similar to growth.

Recommendations

Similar studies between TSR, BPNS, engagement, and growth should be conducted at different levels of schooling as only seventh and eighth grade students were included in this research. While the factors under study in this research did not influence growth, it is possible that other influences come into play at other grade levels. It would also be beneficial to study student growth by subject area, which this research did not do. All subject areas were included and were tested as a single group. It is unknown if multigroup testing would have found the ELA, Math, Science, and Social Studies to be invariant across groups. It is entirely possible that different subjects require different levels of teacher support and engagement and may have influenced student growth percentiles.

The school district and school included in this study have consistently high college and career readiness performance index scores, and students consistently score above the state average on all assessments. The school and district included was also mainly white and affluent, which is different than many schools throughout the state of

Georgia. Similar research should be conducted at schools with varying performance levels, demographic makeup, and level of socioeconomic status throughout the state of Georgia.

Structural equation modeling is a useful statistical technique that can be used to study complex interactions between latent constructs. When identifying latent constructs, it is recommended that latent constructs have at least three indicator variables in the measurement model. The indicators should have a satisfactory reliability coefficient along with high factors loadings showing they adequately reflect the underlying construct. Growth did not fit in the construct of outcome, and if growth will be used in structural equation modeling as an indicator, a better defined construct with growth-like indicators must be formulated. If growth is to be used as an indicator of a latent variable, other indicators similar to growth must be investigated. Since student growth percentiles are a new metric, and none of the investigated indicators or latent constructs had a significant correlation with growth, no recommendation can be made as to other variables that are similar to or influential on student growth percentiles.

To aid in identifying factors that influence student growth percentiles, it would be beneficial to take a more qualitative approach to the research. A strategic sampling method should be implemented, choosing only classrooms with high and low levels of growth. Through classroom observations, using a qualitative approach would allow the researcher to attempt to understand what it is about these teachers or classrooms that lead to student growth or lack thereof. Once common themes have been identified in the strategic samples, researchers could then move into a quantitative research method with random samples to determine if findings are generalizable.

Structural equation modeling is a large sample statistical method with large numbers of participants required for more complex models. Structural equation modeling is an inefficient statistical method to investigate basic relationships between factors. To find other factors that may be related to growth, it is recommended that research initially use simple linear regressions or Pearson correlations to identify variables that influence growth, and then move into the more complex structural equation modeling to study the relationships among constructs.

The more the researcher delved into the results of the questionnaires, the more it seemed that the TSR was not specified as originally intended. Using the previous example, it made logical sense that if I wanted high levels of growth from a student, that student needed to be fully invested and involved in the classroom, teacher, and process. To get students to fully buy into this, there must be a strong teacher-student relationship where teachers have the ability to be role models for their students and get them to do things they normally would not do. While NRI was intended to be used to compare relationships, the researcher believed the results of the NRI did not capture the aspect of the teacher-student relationship that was intended. If the research was to be conducted again, many aspects of the teacher-student relationships similar to Reyes et al., (2012) should be included. Aspects of the TSR would focus on respect for students and their point of view, sensitivity to student needs, genuine interest in students, and warm, caring, and nurturing relationships. The NRI did not pertain to student’s perception of teachers taking a personal interest in students, caring for the students, level of friendship, having respect for students, or being an advocate for the student, which anecdotally, the

researcher has experienced to be characteristics that get students to buy into the classroom and get high levels of growth.

Student growth percentiles can be confusing to understand without some minor investigation. At first glance, it would seem student growth percentiles were just how much students improved from one year to the next, which is just growth in terms of gain scores. Prior research has shown if students are more engaged, there will be better student results. Someone who is uninformed may just assume if they get their students more involved, their growth will increase, which this research contradicted. It would be interesting to see how well educators understand the determination of student growth percentiles and if there was any correlation between teachers who understood how growth was calculated and a level of student growth. Maybe a teacher's understanding of how student growth percentiles are determined can influences student growth.

Conclusion

The purpose of this research was to investigate the impact of teacher-student relationships on student growth percentiles. The research was conducted at a medium-sized middle school that housed seventh and eighth grades students in rural Southwest Georgia. The hypothesized model utilized in the research was based on Self-Determination Theory posited by Deci and Ryan (2009) and the Self-Systems Process Model developed by Connell and Wellborn (1992). The model posited that context (TSR) affects self (BPNS), which affects action (engagement), which consequently affects outcome (outcome). Previously created and documented instruments were used to capture student perceived levels of the teacher-students relationship (Network of Relationships Inventory, Furman and Buhrmester, (2009)), level of psychological need

satisfaction (Psychological Needs Scale, LaGuardia et al.,), and level of classroom engagement (Classroom Engagement Inventory, Wang et al., 2014) at the end of the school year. Student outcomes consisted of GPA, scale scores and norm-referenced scores on the Georgia Milestones assessment, and student growth percentiles calculated by the state of Georgia.

Student responses indicated positive relationships with their teachers, satisfaction of their psychological needs in the classroom, and being engaged in the classroom. Using structural equation modeling, the relationship between constructs and indicators was investigated with much of the prior research supported by the current findings. The research concluded that the Self-systems Process Model posited by Connell and Wellborn (1991) was plausible as there was good fit to the data. In both structural equation models with scale score and growth, positive teacher-student relationships had an impact on BPNS, engagement, and student outcomes. A positive TSR had a direct effect on engagement and an indirect effect, acting through the mediating variable, BPNS. TSR, while not having a statistically significant direct effect on outcome, had an indirect effect acting through the mediating variable, engagement. Engagement had a small but statistically significant impact on student outcomes.

In the structural model that included growth as an indicator of outcome, it was determined that growth did not fit in the model. Growth had a low factor loading and was not statistically significant. While TSR had positive association with BPNS and engagement, and engagement was positively associated with student outcomes, there was no association with a positive TSR and student growth. The findings were consistent when comparing LowSES and HighSES groups and comparing White and NonWhite

groups. So while the models demonstrated that teachers that foster a positive TSR will lead to better student outcomes when defined by NormRef, ScScr, and GPA, those same positive relationships show no influence on student growth.

Concluding Thoughts

Growth is a new metric that has not been well defined statistically, either in this research or any of the literature in the review. Student growth now plays a significant role in teacher evaluations in the state of Georgia, and identifying strategies teachers can implement in their classrooms to ensure student growth has becoming increasingly important. Indicators and constructs included in this research showed little to no correlation with student growth, which is contrary to what the researcher expected.

With the way SGPs are determined and my own classroom experiences, I believed that a positive TSR would result in positive growth. Working backwards from outcome, high growth can be achieved by a student by getting that student to perform better on the current assessment compared to his or her academic peers throughout the state. The way to get the student to perform better is to prepare him or her better for the assessment than his or her academic peers. That would include getting the student to willingly participate in all the classroom activities designed to prepare for the assessment. The student would need to be on task and genuinely trying to do their best on assigned activities, not only for themselves, but also for the benefit of the teacher-student relationship. Having a positive TSR and satisfying BPNS should foster student

engagement and therefore have a positive impact on student growth; however, did not happen.

Future research should investigate other factors that may influence student growth using more efficient statistical methods. There is still little evidence, other than the findings of a dissertation completed by LeGeros (2013), of factors that influence student growth percentiles. These findings are concerning since many evaluation systems in use throughout the United States, utilized student growth percentiles as a portion of teacher evaluation, which should drive future research.

REFERENCES

Archambault, I., Janosz, M., Fallu, J., & Pagani, L. S. (2009). Student engagement and its relationship with early high school dropout. *Journal of adolescence*, *32*(3), 651-670.

Bagozzi, R. P., & Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the academy of marketing science*, *40*(1), 8-34.

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. Macoulides & R. Schumacker (Eds), *New developments and techniques in structural equation modeling* (pp. 269-296). Mahwah: Lawrence Erlbaum Associates.

Barnett, J. H., & Amrein-Beardsley, A. (2011). *Actions over credentials: Moving from highly qualified to measurably effective* [Commentary]. Teachers College Record. Retrieved from http://www.tcrecord.org.proxy-remote.galib.uga.edu/Content.asp?ContentID=16517

Batista, I. A. (2014). *A comparison of a value added status model versus a value added growth model for identifying high performing Maine middle schools* (Master's thesis). Retrieved from http://digitalcommons.usm.maine.edu/muskie_capstones/84/

Bergkvist, L. (2015). Appropriate use of single-item measures is here to stay. *Marketing Letters*, *26*(3), 245-255.

Betebenner, D. (2008). *Norm- and criterion-referenced student growth.* The National Center for the Improvement of Educational Assessment. Retrieved January 13, 2015 from http://www.nciea.org/publications/normative_criterion_growth_DB08.pdf

Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, *28*(4), 42–51.

Betebenner, D. (2011). *A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth trajectories/projections*. The National Center for the Improvement of Educational Assessment. Retrieved Januеary 20, 2015 from http://www.nj.gov/education/njsmart/performance/SGP_Technical_Overview.pdf

Birch, S. H., & Ladd, G. W. (1997). The teacher-child relationship and children's early school adjustment. *Journal of school psychology*, *35*(1), 61-79.

Birch, S. H., & Ladd, G. W. (1998). Children's interpersonal behaviors and the teacher–child relationship. *Developmental psychology*, *34*(5), 934.

Briggs, D. C., Dadey, N., & Kizil, R.C. (2014a). *Adjusting mean growth percentiles for classroom composition*. University of Colorado.

Briggs, D. C., Dadey, N., & Kizil, R. C. (2014b). *Comparing student growth and teacher observation to principal judgments in the evaluation of teacher effectiveness*. University of Colorado.

Brown, T. A., & Moore, M. T. (2015). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361-379). New York, NY: Guilford Press.

Buhrmester, D. & Furman, W. (2008). *The network of relationships inventory: Relationship qualities version*. Unpublished measure, University of Texas at Dallas.

Burniske, J., & Melbaum, D. (2012). *The use of student perceptual data as a measure of teaching effectiveness: Texas Comprehensive Center Briefing Paper, Number 8*. Retrieved from Advancing Research Improving Education website: http://www.sedl.org/pubs/catalog/items/txcc08.html

Buzick, H. M., and Laitusis, C. C. (2010). *A summary of models and standards-based applications for grade-to-grade growth on statewide assessments and implications for students with disabilities* (Educational Testing Service ETS RR-10-14). Princeton, NJ: ETS. Retrieved from http://www.ets.org/Media/Research/pdf/RR-10-14.pdf

Bylsma, P. J. (2014). Using SGPs to measure student growth: Context, characteristics, and cautions. *The WERA Educational Journal*, *6*(1), 10-19.

Byrne, B. M. (2004). Testing for multigroup invariance using AMOS graphics: A road less traveled. *Structural Equation Modeling*, *11*(2), 272-300.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York, NY: Routledge.

Castellano, K. E., & Ho, A. D. (2013). *A practitioner's guide to growth models*. Washington, DC: CCSSO. Retrieved from http://scholar.harvard.edu /files/andrewho/files/a_pracitioners_guide_to_growth_models.pdf

Cervoni, J. M. (2014). *Factors that influence teacher growth scores* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3582095)

Collins, C., Amrein-Beardsley, A. (2014). Putting growth and value added models on the map: A national overview. *Teachers College Record*, *116*(1), 1-32.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In R. Gunnar & L.A. Sroufe (Eds.), *Self processes in development: Minnesota symposium on child psychology*, (pp. 43-77). Chicago: Chicago University Press.

Conner, J. O., & Pope, D. C. (2013). Not just robo-students: Why full engagement matters and how schools can promote it. *Journal of youth and adolescence*, *42*(9), 1426-1442.

Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research, 77*(1), 113-143.

Craig, T. A. (2011). *Effects of standards-based report cards on student learning* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3498282)

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation?. *Educational Researcher, 44*(2), 132-137.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Deci, E. L., & Ryan, R. M. (2000a). The darker and brighter sides of human existence: Basic Psychological needs as a unifying concept. *Psychological Inquiry*, *11*(4), 319-338.

Deci, E. L., & Ryan, R. M. (2000b). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, *11*(4), 227-268.

Deci, E.L., & Ryan, R.M. (2002). Self-determination research: Reflections and future directions. In E.L. Deci & R.M. Ryan (Eds.), *Handbook of self-determination research* (pp. 431–441). Rochester, NY: University of Rochester Press.

Deci, E. L., & Ryan, R. M. (2009). Self-determination theory: a consideration of

human motivational universals. In P. J. Corr & G. Matthews (Eds.), *The Cambridge handbook of personality psychology* (pp. 441-455). New York, NY: Cambridge University Press.

Dollan, C. V. (1994) Factor analysis of variables with 2, 3, 4, and 7 response categories: A comparison of categorical variables estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47*(2), 309–326.

Doran, H. C. (2003). The challenges of accountability: Adding value to accountability. *Educational Leadership. 61*(3), 55-59.

Duffield, S., Wageman, J., & Hodge, A. (2013). Examining how professional development impacted teachers and students of US history courses. *The Journal of Social Studies Research*, *37*(2), 85-96.

Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. (2012). Selecting growth measures for school and teacher evaluations. Working Paper 80. *National Center for Analysis of Longitudinal Data in Education Research*.

Fisher, D., & Rickards, T. (1998). Associations between teacher-student interpersonal behaviour and student attitude to mathematics. *Mathematics Education Research Journal*, *10*(1), 3-15.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59-109.

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763-782). New York, NY: Springer.

Fried, L. J., & Konza, D. M. (2013). Using self-determination theory to investigate student engagement in the classroom. *The International Journal of Pedagogy and Curriculum, 19*(2), 25-41.

Furman, W., & Buhrmester, D. (1985). Children's perceptions of the qualities of sibling relationships. *Child development*, *56*(2), 448-461.

Furman, W., & Buhrmester, D. (1985). Children's perceptions of the personal relationships in their social networks. *Developmental psychology*, *21*(6), 1016-1024.

Furman, W., & Buhrmester, D. (2009). *Network of relationships questionnaire manual*. Unpublished manual, University of Denver, Colorado.

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, *95*(1), 148-162.

Gabriel, T. (2010, September 2). A celebratory road trip for education secretary. *New York Times*, Retrieved from http://www.nytimes.com/2010/09/02/education/02duncan.html

Garn, A. C., & Wallhead, T. (2015). Social goals and basic psychological needs in high school physical education. *Sport, Exercise, and Performance Psychology*, *4*(2), 88-99.

Garson, G. D. (2012). *Testing statistical assumptions*. Asheboro, NC: Statistical Publishing Associates.

Garson, G. D. (2015). *Missing values analysis & data imputation.* Asheboro, NC : Statistical Publishing Associates.

Georgia Department of Education, Curriculum, Instruction and Assessment. (2013). *Parents' Guide to New Tests in Georgia*. Retrieved from http://www.pta.org/files/Advocacy/CCSSIToolkit/Common%20Core%20State%20Standards%20Resources/Assessments%20Resouces/PTA_GA_6PG_17DEC13_FINAL.pdf

Georgia Department of Education, Curriculum, Instruction and Assessment. (2014a). *Methods of combining SGPs*. Retrieved from http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Documents/Methods%20of%20Combining%20SGPs.pdf

Georgia Department of Education, Curriculum, Instruction and Assessment. (2014b). *Overview of the Georgia student growth model*. Retrieved from http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Documents/GSGM%20Overview.pdf

Georgia Department of Education, Office of School Improvement, Teacher and Leader Effectiveness Division. (2014a). *Teacher keys effectiveness system*. Retrieved from http://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Documents/FY15%20TKES%20and%20LKES%20Documents/TKES%20Handbook%20-%20FINAL%2010-15-14.pdf

Georgia Department of Education, Office of School Improvement, Teacher and Leader Effectiveness Division. (2014b). *TEM scoring guide and methodology*. Retrieved from http://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Documents/TEM%20Scoring%20Guide%206-18-14Final1.pdf

Goldhaber, D., Walch, J., & Gabele, B. (2014). Does the model matter? Exploring the

relationship between different student achievement-based teacher assessments. *Statistics and Public Policy, 1*(1), 28-39.

Goldschmidt, P., Roschewski, P., Choi, K. C., Auty, W., Hebbler, S., & Williams, A. (2005). *Policymakers' guide to growth models for school accountability: How do accountability models differ?* Washington, DC: CCSSO. Retrieved from http://www.ccsso.org/Documents/2005/Policymakers_Guide_To_Growth_2005.pdf

Guarino, C., Reckase, M., Stacy, B., Wooldridge, J. (2014). A comparison of growth percentile and value added models of teacher performance (Working Paper #39). Michigan State University: The Education Policy Center at Michigan State University. Retrieved from http://education.msu.edu/epc/publications/documents/WP39AComparisonofGrowthPercentileandvalue addedModel.pdf

Haertel, E. H. (2013). *Reliability and validity of inferences about teachers based on student test scores*. Princeton, NJ: Educational Testing Service.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher–child relationships and the trajectory of children's school outcomes through eighth grade. *Child development*, *72*(2), 625-638.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London & New York: Routledge.

Henderson, D. (1995, April). *Associations between learning environments and student outcomes in biology*. Paper presented at the Annual Meeting of American Educational Research Association, San Francisco, CA. Retrieved from http://files.eric.ed.gov/fulltext/ED390704.pdf

Holgado–Tello, F. P., Chacón–Moscoso, S., Barbero–García, I., & Vila–Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, *44*(1), 153-166.

Hoyle, R. H. (2015). *Handbook of structural equation modeling*. New York, NY: Guilford Press.

Huitt, W., Huitt, M., Monetti, D., & Hummel, J. (2009). *A systems-based synthesis of research related to improving students' academic performance*. Paper presented at the 3rd International City Break Conference sponsored by the Athens Institute for Education and Research (ATINER), October 16-19, Athens, Greece. Retrieved from http://www.edpsycinteractive.org/papers/improving-school-achievement.pdf

Hughes, J. N., Luo, W., Kwok, O., & Loyd, L. K. (2008). Teacher-student support, effortful engagement, and achievement: A 3-year longitudinal study. *Journal of Educational Psychology*, *100*(1), 1-14.

Hughes, J. N., Wu, J., Kwok, O., Villarreal, V., & Johnson, A. Y. (2012). Indirect effects of child reports of teacher–student relationship on achievement. *Journal of Educational Psychology*, *104*(2), 350-365.

In'nami, Y., & Koizumi, R. (2013). Structural equation modeling in educational research: A primer. In M. S. Khine (Ed.), *Application of Structural Equation Modeling in Educational Research and Practice* (pp. 23-51). Boston: Sense Publishers.

Joshua, M. T., Joshua, A. M., & Kritsonis, W. A. (2006). Use of student achievement scores as basis for assessing teachers' instructional effectiveness: Issues and research results. *National Forum Teacher Education Journal, 16*(3), 1-13.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, 4th ed., pp. 233-265). Boston, MA: McGraw-Hill.

Klem, A. M., & Connell, J. P. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of School Health*, *74*(7), 262-273.

Kline, R. B. (2011). *Principles and practices of structural equation modeling*. New York: The Guilford Press.

Ladd, H. F., & Lauen, D. L. (2010). Status versus growth: The distributional effects of school accountability policies. *Journal of Policy Analysis and Management*, *29*(3), 426-450.

Larwin, K., & Harvey, M. (2012). A demonstration of a systematic item-reduction approach using structural equation modeling. *Practical Assessment, Research & Evaluation*, *17*(8), 1-19.

La Guardia, J. G., Ryan, R. M., Couchman, C. E., & Deci, E. L. (2000). Within-person variation in security of attachment: a self-determination theory perspective on attachment, need fulfillment, and well-being. *Journal of personality and social psychology*, *79*(3), 367-384.

LeGeros, L. (2013). *The association between elementary teacher licensure test scores and student growth in mathematics: An analysis of Massachusetts MTEL and MCAS tests* (Doctoral dissertation). Retrieved from http://scholarworks.umb.edu/

Lei, P. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality and Quantity*, *43*(3), 495-507.

Lawson, M. A., & Lawson, H. A. (2013). New conceptual frameworks for student

engagement research, policy, and practice. *Review of Educational Research*, *83*(3), 432-479.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*(2), 151-173.

Mahatmya, D., Lohman, B. J., Matjasko, J. L., & Farb, A. F. (2012). Engagement across developmental periods. In S.L. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 45-63). New York, NY: Springer.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal*, *37*(1), 153-184.

Martin, A. J. (2014). Interpersonal Relationships and Students' Academic and Non-Academic Development. In D. Zandvliet, P. den Brok, T. Mainhard, & J. van Tartwijk (Eds.), *Interpersonal relationships in education: From theory to practice* (pp. 9-24). Boston, MA: Sense Publishers.

McCaffrey, D. F., & Castellano, K. E. (2014). *A review of comparisons of aggregated student growth percentiles and value added for educator performance measurement*. Princeton, NJ: Educational Testing Service.

Morata-Ramirez, M. A., & Holgado-Tello, F. P. (2013). Construct validity of Likert scales through confirmatory factor analysis: A simulation study comparing different methods of estimation based on pearson and polychoric correlations. *International Journal of Social Science Studies, 1*(1), 54-61.

Muthén, B., & Asparouhov, T. (2011). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313-335.

Nachtigall, C., Kroehne, U., Funke, F., & Steyer, R. (2003). Should we use SEM? Pros and cons of structural equation modeling. *Methods of Psychological Research Online*, *8*(2), 1-22.

Newsom, J. (2005). *Practical approaches to dealing with nonnormal and categorical variables* [PDF Document]. Retrieved from Lecture Notes Online Web site: http://www.upa.pdx.edu/IOA/newsom/semclass/

Nichols, S. L., Glass, G. V., & Berliner, D. C. (2005). *High-Stakes testing and student achievement: Problems for the No Child Left Behind Act* (EPSL-0509-105-EPRU). Tempe, AZ: Arizona State University.

Ntoumanis, N. (2005). A prospective study of participation in optional school physical education using a self-determination theory framework. *Journal of Educational*

Psychology, *97*(3), 444-453.

O'Malley, K. J., Murphy, S., McClarty, K. L., Murphy, D., & McBride, Y. (2011). *Overview of student growth models* (White Paper). Retrieved from Pearson website: http://www.pearsonassessments.com/hai/Images/tmrs/Student_Growth_WP_083111_FINAL.pdf

Opdenakker, M., & Minnaert, A. (2014). Learning environment experiences in primary education. In D. Zandvliet, P. den Brok, T. Mainhard, & J. van Tartwijk (Eds.), *Interpersonal relationships in education: From theory to practice* (pp. 183-194). Boston, MA: Sense Publishers.

Osunsami, S., & Forer, B. (2011, July 6). Atlanta cheating: 178 teachers and administrators changed answers to increase test scores. *ABC News*. Retrieved from http://abcnews.go.com/US/atlanta-cheating-178-teachers-administrators-changed-answers-increase/story?id=14013113

Petrescu, M. (2013). Marketing research using single-item indicators in structural equation models. *Journal of Marketing Analytics, 1*(2), 99-117.

Prince, C. D. Schuermann, P. J., Guthrie, J. W., Witham, P. J., Milanowski, A. T., & Thorn, C. A. (2009). The other 69 percent: Fairly rewarding the performance of teachers of non-tested subjects and grades. Washington, DC: Center for Educator Compensation Reform.

Pianta, R. C. (2001). *STRS: Student-teacher Relationship Scale: Professional manual*. Psychological Assessment Resources.

Pianta, R. C., Hamre, B. K., & Allen, J. P. (2012). Teacher-student relationships and engagement: Conceptualizing, measuring, and improving the capacity of classroom interactions. In S.L. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 365-386). New York, NY: Springer.

Reddy, R., Rhodes, J. E., & Mulhall, P. (2003). The influence of teacher support on student adjustment in the middle school years: A latent growth curve study. *Development and Psychopathology*, *15*(1), 119-138.

Reeve, J. (2012). A self-determination theory perspective on student engagement. In S. L. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 149-172). New York, NY: Springer.

Reschly, A. L., & Christenson, S. L. (2006). Prediction of dropout among students with mild disabilities: A case for inclusion of student engagement variables. *Remedial and Special Education, 27*(5), 276–292.

Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In S.L. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 3-19). New York, NY: Springer.

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology*, *104*(3), 700-712.

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354-373.

Rickards, T., & Fisher, D. L. (1997, July). *A report of research into student attitude and teacher student interpersonal behaviour in a large sample of Australian secondary mathematics classrooms*. Paper presented at the meeting of the Mathematics Education Research Group of Australia, Rotorua, New Zealand.

Rigdon, E. (1997, June). Not positive definite matrices - Causes and cures. Retrieved from http://www2.gsu.edu/~mkteer/npdmatri.html

Roorda, D., Koomen, H., Split, J. L., & Oort, F. J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research, 81*(4), 493-529.

Rudasill, K. M., & Rimm-Kaufman, S. E. (2009). Teacher–child relationship quality: The roles of child temperament and teacher–child interactions. *Early Childhood Research Quarterly*, *24*(2), 107-120.

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, *52*(1), 141-166.

Ryser, G. R., & Rambo-Hernandez, K. E. (2014). Using growth models to measure school performance. *Gifted Child Today*, *37*(1), 17-23.

Sakiz, G., Pape, S. J., & Hoy, A. W. (2012). Does perceived teacher affective support matter for middle school students in mathematics classrooms?. *Journal of School Psychology, 50*(2), 235-255.

Satorra, A., & Bentler. (1994). Corrections to test statistics and standard errors in

covariance structure analysis. In A. V. Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications to developmental research* (pp. 399-419). Thousand Oaks, CA: SAGE Publications Inc.

Schafer, W. D., Lissitz, R. W., Zhu, Z., & Zhang, Y. (2012). Evaluating teachers and schools using student growth models. *Practical Assessment, Research & Evaluation, 17*(17), 1-12.

Schiro, M. S. (2013). *Curriculum theory: Conflicting visions and enduring concerns*. Thousand Oaks, California: Sage Publications, Inc.

Schochet, P. Z., & Chiang, H. S. (2010). *Error rates in measuring teacher and school performance based on student test score gains* (NCEE 2010-4004). National Center for Education Evaluation and Regional Assistance.

Sever, M., Ulubey, Ö., Toraman, Ç., & Türe, E. (2014). An analysis of high school students' classroom engagement in relation to various variables. *Education and Science, 39*(176), 183-198.

Simmons, D. R. (2006). *The relationship between seven teacher as a person traits and student growth on the Idaho standard achievement test* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3209106)

Skinner, E. A., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic?. *Journal of Educational Psychology*, *100*(4), 765-781.

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009) A motivational perspective on engagement and disaffection: conceptualization and assessment of children’s behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69*(3), 493-525.

Skinner, E. A., & Pitzer, J. R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience. In S.L. Christenson, A. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 21-44). New York, NY: Springer.

Smart, J. B. (2014). A mixed methods study of the relationship between student perceptions of teacher-student interactions and motivation in middle level science. *RMLE Online*, *38*(4), 1-19.

Song, X., & Lee, S. (2012). A tutorial on the Bayesian approach for analyzing structural equation models. *Journal of Mathematical Psychology, 56*(3), 135-148.

Stroet, K., Opdenakker, M., & Minnaert, A. (2013). Effects of need supportive teaching on early adolescents’ motivation and engagement: A review of the literature. *Educational Research Review, 9*, 65-87.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. (6th ed.). New York, NY: Pearson.

Templin, J. (2012). *Measurement invariance* [PDF document]. Retrieved from Lecture Notes Online Web site: http://jonathantemplin.com/files/sem/sem12ersh8750/sem12ersh8750_lecture11.pdf

Teo, T., Tsai, L. T., & Yang, C. C. (2013). Applying structural equation modeling (SEM) in educational research: An introduction. In M. S. Khine (Ed.), *Application of structural equation modeling in educational research and practice* (pp. 3-21). Boston: Sense Publishers.

The American Heritage College Dictionary (3rd ed.). (1993). New York: Houghton Mifflin Co.

Thurlow, M., Lazarus, S., Quenemoen, R., & Moen, R. (2010). *Using growth for accountability: Considerations for students with disabilities* (NCEO Policy Directions – Number 21). Minneapolis, MN: National Center on Educational Outcomes.

Tian, L., Han, M., & Huebner, E. S. (2014). Preliminary development of the Adolescent Students' Basic Psychological Needs at School Scale. *Journal of adolescence, 37*(3), 257-267.

United States Department of Education. (2009). *Race to the top program.* Washington, DC: Author. Retrieved from https://www2.ed.gov/programs/racetothetop/executive-summary.pdf

Unite States Department of Education (2011a). Final report on the evaluation of the growth model pilot project, Washington, D.C: Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. Retrieved from http://www2.ed.gov/rschstat/eval/disadv/growth-model-pilot/gmpp-final.pdf

Wang, Z., Bergin, C., & Bergin, D. A. (2014). Measuring engagement in fourth to twelfth grade classrooms: The classroom engagement inventory. *School Psychology Quarterly, 29*(4), 517-535.

Wentzel, K. R. (2002). Are effective teachers like good parents? Teaching styles and student adjustment in early adolescence. *Child Development, 73*(1), 287-301.

Wilkins, J. (2014). Good teacher-student relationships: Perspectives of teachers in urban high schools. *American Secondary Education, 43*(1), 52-68.

Wonglorsaichon, B., Wongwanich, S., & Wiratchai, N. (2014). The influence of

students school engagement on learning achievement: A structural equation modeling analysis. *Procedia-Social and Behavioral Sciences*, *116*(21), 1748-1755.

Wothke, W. (1993). Nonpositive definite matrices in structural equation modeling, In K.A. Bollen & J.S. Long (Eds), *Testing structural equation models* (pp. 256-293). Newbury Park, CA: SAGE Publications

Wubbels, T. & Levy, J. (1993). *Do you know what you look like: Interpersonal relationships in education*, London, UK: The Falmer Press.

Wyse, A. E., & Dong, G. S. (2014). A comparison of three conditional growth percentile methods: Student growth percentiles, percentile rank residuals, and a matching method. *Practical Assessment, Research & Evaluation*, *19*(15), 1-12.

Yazzie-Mintz, E. (2010). *Charting the path from engagement to achievement*. Bloomington, IN: Center for Evaluation and Education Policy.

Appendix A

Columbus State University IRB Approval

CSU IRB 4:35 PM (3 hours ago)

to me, Deirdre, HardinS, Amber, Clayton, cotten_brett, David, Diana, Ellen, Gregory, Iris, Je

Institutional Review Board
Columbus State University

Date: 3/2/16
Protocol Number: 16-064
Protocol Title: The Impact of Teacher-Student Relationships and Classroom Engagement on Student Growth Percentiles of 7th and 8th Grade Students in One Rural School in Southwest Georgia
Principal Investigator: David Dennie
Co-Principal Investigator: Deirdre Greer

Dear David Dennie:

Representatives of the Columbus State University Institutional Review Board have reviewed your research proposal identified above. It has been determined that the research project poses minimal risk to subjects and qualifies for expedited review under 45 CFR 46.110.

Approval is granted for one (1) year from the date of this letter for approximately 800 subjects. Please note any changes to the protocol must be submitted in writing to the IRB before implementing the change(s). Any adverse events, unexpected problems, and/or incidents that involve risks to participants and/or others must be reported to the Institutional Review Board at irb@columbusstate.edu or (706) 507-8634.

You must submit a Final Report Form to the IRB once the project is completed or within 12 months from the date of this letter. If the study extends beyond 1 year, you must submit a Project Continuation Form to the IRB. Both forms are located on the CSU IRB website (http://research.columbusstate.edu/irb/). The completed form should be submitted to irb@columbusstate.edu. Please note that either the Principal Investigator or Co-Principal Investigator can complete and submit this form to the IRB. Failure to submit this required form could delay the approval process for future IRB applications.

If you have further questions, please feel free to contact the IRB.

Sincerely,
Amber Dees, IRB Coordinator

Institutional Review Board
Columbus State University

Appendix B

Superintendent Letter of Permission

[redacted]

[redacted]

[redacted]

1/21/2016

David A Dennie
1536 Goat Rock Road
Fortson, GA 31808

Dear David Dennie:

Based on my review of your proposed research project, I grant permission for you to conduct the study entitled "The impact of teacher-student relationships and classroom engagement on student growth percentiles of seventh and eighth grade students in or rural school in southwest Georgia" within the [redacted], specifically, [redacted]. As part of this study, I authorize you to recruit all middle school students to participate in the study, have students complete the inventory of classroom relationships, psychological needs satisfaction, and engagement. You will also have access to student status and growth scores, provided by the curriculum department for the 2015-2016 school year to be combined with student inventory data. At no point should any school, teacher, and/or student information be published. I understand that the resources required for this research project include student time and computer lab use.

Sincerely,

[redacted]

Appendix C

Principal Letter of Permission

[redacted]

[redacted]

1/21/2016

David A Dennie
1536 Goat Rock Road
Fortson, GA 31808

Dear David Dennie:

Based on my review of your proposed research project, I grant permission for you to conduct the study entitled "The impact of teacher-student relationships and classroom engagement on student growth percentiles of seventh and eighth grade students in or rural school in southwest Georgia" at [redacted] As part of this study, I authorize you to recruit all middle school students to participate in the study, have students complete the inventory of classroom relationships, psychological needs satisfaction, and engagement. You will also have access to student status and growth scores, provided by the curriculum department for the 2015-2016 school year to be combined with student inventory data. At no point should any school, teacher, and/or student information be published. I understand that the resources required for this research project include student time and computer lab use.

Sincerely,

[redacted]

Appendix D

Parental Informed Consent Form

PARENTAL INFORMED CONSENT FORM
COLUMBUS STATE UNIVERSITY

Your son/daughter is being asked to participate in a research project conducted by Dave Dennie, a doctoral student in the Curriculum and Leadership program at Columbus State University. The research project will be supervised by Dr. Deidre Greer, Dean of the College of Education and Health Professions at Columbus State University and will take place at [redacted].

I. Purpose:
The purpose of this project is to identify how student perceived relationships between teachers and students and a student's level of engagement in the classroom impact status and growth scores on the Georgia Milestones assessment. The results of this study should help to further educators understanding of classroom that impact student achievement and growth on the Georgia Milestones assessments.

II. Procedures:
Your son/daughter will complete a 46 question inventory pertaining to their perception of the teacher-student relationship, psychological needs satisfaction, and levels of engagement in the classroom. Within the inventory, students will also be asked to identify their classroom teacher, subject area, grade level, and their student ID. The classroom teacher will NOT be present while students are completing the inventory. The inventory will take approximately thirty minutes to complete in the spring of 2016 at the discretion of the school principal. All student responses will be strictly confidential and will NOT be view by any school personnel.

III. Possible Risks or Discomforts:
Student ID’s will be used to merge individual state assessment results and student inventory responses. Students may feel that teachers will see their personal responses which could make them feel uncomfortable with responding to the inventory. Teachers will not be present during inventory administration and will not have access to student responses. Any information students provide will be kept confidential on the researchers

personal computer which is password protected. No person other than the primary researcher will have access to student responses.

IV. Potential Benefits:

There is currently no research that identifies factors that can aid student growth as determined by student growth percentiles with the classroom. This investigation will attempt to determine if there is a link between teacher-student relationships, psychological need satisfaction, student engagement, and student growth percentiles. If a positive link is found between any of the investigated factors and student growth percentiles, the information could be used to improve can be used to aid educators in improving student growth and overall student achievement levels.

V. Costs and Compensation:

There will be no cost or compensation for your child’s participation in this research study.

VI. Confidentiality

The 46 question inventory will be completed electronically using Google Forms under the researchers personal Google account and the data will be inaccessible by others. While student ID’s will be part of the survey, no one except the researcher will see recorded results. Status and growth scores will be collected from the [redacted] curriculum department and merged with the inventory data under the same Google account using student ID’s.

Upon completion of data analysis and successful dissertation defense, student ID's will be removed from the data file and the file will be retained for one year. Following the completion of the dissertation, the possibility exists that findings will be published. Your child’s individual privacy will be maintained in all published and written data resulting from the study. Names of students, teachers, and school will NOT be used in the published manuscript and all information will be kept strictly confidential.

VII. Withdrawal:

Your child’s participation in this study is completely voluntary. He/she will also be given the option of participating. Both his/her assent (agreement) and your permission are required for him/her to be a part of this study. Your child may elect to withdraw from this study at any time and the information I have collected will be destroyed. Refusal to participate or withdrawal from the study at any time will involve no penalty or loss of benefits.

For additional information about this research project, you may contact the principle investigator, Dave Dennie at [redacted] or [redacted]. If you have questions about your rights or your son/daughters as a research participant, you

may contact Columbus Status University Institutional Review Board at irb@columbusstate.edu.

I have read this parental informed consent form provided to me. If I had any questions, they have been answered to my satisfaction. By signing this form, I agree to allow my child to participate in this study. Parents choosing to allow their son/daughter to participate in the research students will sign the informed consent document and have their son/daughter return the document to the front office of [redacted].

____

Printed Student Name

____

Printed Parent/Guardian Name

____ ____________

Parent/Guardian Signature Date

Appendix E

Student Assent Form

STUDENT ASSENT TO PARTICIPATE IN RESEARCH STUDY

THE IMPACT OF TEACHER-STUDENT RELATIONSHIPS AND CLASSROOM ENGAGEMENT ON STUDENT GROWTH PERCENTILES OF 7TH AND 8TH GRADE STUDENTS IN ONE RURAL SCHOOL IN SOUTHWEST GEORGIA

My name is Dave Dennie and I am a doctoral student in the Curriculum and Leadership program at Columbus State University. I am doing a research study and would like to tell you about this study and ask if you will take part (be a "subject") in it.

What is a research study?

A research study is when people like me collect a lot of information about a certain thing to find out more about it. Before you decide if you want to be in this study, it's important for you to understand why I am doing the research and what is involved. Please read this form carefully. You can discuss it with your parents or anyone else. If you have questions about this research, you can email me at [redacted] or ask me prior to starting the research survey.

Why are we doing this study?

The purpose of this project is to identify how student perceived relationships between teachers and students and a student's level of engagement in the classroom impact status and growth scores on the Georgia Milestones assessment. The results of this study should help to further educators understanding of classroom factors that impact student achievement and growth on the Georgia Milestones assessments.

What will happen if you are in this study?

If you agree to participate in this study, you will be asked to complete a 46-question survey pertaining to your perception of your teacher-student relationship, your basic psychological needs satisfaction, and your level of engagement in the classroom. The survey is asking what YOU believe to be true and not what other students and teachers believe. There are no right or wrong answers and your responses indicate what is happening in your classroom. The survey will take about 30 minutes to complete.

<u>Who will know about your study participation?</u>

As the researcher, I will be the only one who will know the details of your study participation. No one else will have access to your responses. Within the survey, along with the 46-question survey, you will also be asked to identify your classroom teacher, subject area, grade level, and student ID. Your teacher will not be present during survey administration and will not have access to your responses.

Following the completion of the research study, the possibility exists that findings will be published. I will not use your name, your teacher's name, your school's name or any other personal information that would identify you in any published and written data resulting from the study. All information will be kept strictly confidential. Upon completion of data analysis and successful dissertation defense, student ID's will be removed from the data file and the file will be retained for one year.

<u>Are there any risks or discomforts to being in the study</u>?

Your student ID will be used to merge your individual state assessment results with your survey responses. You may feel that your teachers will see your responses, which could make you feel uncomfortable with responding honestly on the survey. Teachers will not be present during survey administration and will not have access to your responses. Any information you provide will be kept confidential on the researchers personal computer, which is password protected. No person other than the primary researcher will have access to your responses.

<u>Are there any benefits to being in the study?</u>

This investigation will attempt to determine if there is a link between teacher-student relationships, psychological need satisfaction, student engagement, and student growth percentiles. If a positive link is found between any of the investigated factors and student growth percentiles, the information could be used to aid educators in improving student growth and overall student achievement levels.

<u>Will you get paid for being in the study</u>?

There will be no cost or compensation for your child's participation in this research study.

<u>Do you have to be in the study</u>?

You do not have to participate in the research study. Research is something you do only if you want to and your participation in this study is completely voluntary. No one at the school will get mad at you if you do not want to be in the study, and refusal to participate or withdrawal from the study at any time will involve no penalty, loss of benefits, or have an impact on your grades and standing in your class. You can quit at any time.

Do you have any questions?
You can contact the researcher at [redacted] if you have questions about the study or ask questions prior to survey administration.

Student Assent

If you decide to participate, and your parents agree, we'll give you a copy of this form to keep for future reference.

If you would like to be in this research study, please print and sign your name on the line below.

____ _______________
Child's Printed Name Date

____ _______________
Child's Signature Date

____ _______________
Signature of Investigator/Person Obtaining Assent Date

Appendix F

Permission to use Network of Relationships Inventory

Re: Network of Relationships Inventory - Relationship Quality Version Inbox x

Wyndol Furman <Wyndol.Furman@du.edu> Aug 24 (4 days ago)
to me

Yes, you have permission.

Dr. Wyndol Furman
John Evans Professor and Director of Clinical Training
Department of Psychology
University of Denver
Denver, CO 80208
(e) wfurman@nova.psy.du.edu
(p) 303-871-3688
(f(303-871-4747
http://www.du.edu/psychology/relationshipcenter/

From: David Dennie <dennie.david@gmail.com>
Date: Monday, August 24, 2015 at 6:53 PM
To: Wyndol Furman <Wyndol.Furman@du.edu>
Cc: David Dennie <dennie.david@gmail.com>
Subject: Network of Relationships Inventory - Relationship Quality Version

Dr. Furman, my name is Dave Dennie and I am a doctoral student at Columbus State University in Columbus Georgia. I am in the process of developing my dissertation proposal on teacher-student relationships, engagement, and student growth percentiles in Georgia and would like to request permission to utilize the NRI-RQV in my dissertation as a measure of the teacher-student relationship. According to the manual, permission is given to copy and use the scale, however I just wanted to verify that I could use the instrument in my research. Can you please respond letting me know that it is approved for me to use the NRI-RQV? Thank you for any help you can provide and taking the time to read and respond to my email.

Dave Dennie

Appendix G

Permission to use Basic Psychological Needs Inventory

Jennifer La Guardia Aug 20 (1 day ago)
to me

Dave,

You are welcome to use the scale...it is in the public domain and does not require permission for its use.

The main article with all of the validation information is:
La Guardia, J. G., Ryan, R. M., Couchman, C. E., & Deci, E. L. (2000). Within-person variation in security of attachment: A self-determination theory perspective on attachment, need fulfillment, and well-being. *Journal of Personality and Social Psychology*, *79*, 367-384.

You can find a downloadable version on the SDT website, along with information on how to calculate the scale and describe it.

Hope that helps.

Good luck with your research.

Jennifer La Guardia

On 8/19/2015 3:29 PM, David Dennie wrote:

> Dr. La Guardia, my name is Dave Dennie and I am a doctoral student at Columbus State University in Columbus Georgia. I am in the process of developing my dissertation proposal on teacher-student relationships, basic psychological needs, engagement, and student growth percentiles in Georgia and would like to formally request permission to utilize the need satisfaction scale developed in the article Within-Person Variation in Security of Attachment: A Self-Determination Theory Perspective on Attachment, Need Fulfillment, and Well-Being in my dissertation. Can you please respond letting me know that is approved for me to use the Needs Satisfaction Scale? Also, if you could, I would appreciate any other supporting documentation that you may have from the development of the scale.
>
> Thank you for any help you can provide and taking the time to read and respond to my email. I look forward to hearing from you.
>
> Dave Dennie
> Columbus State University
> Doctoral Candidate

Appendix H

Modified Basic Psychological Needs Inventory

In My Relationships

1	2	3	4	5	6	7
not at all true			somewhat true			very true

1. When I am with my teacher, I feel free to be who I am.

2. When I am with my teacher, I feel like a competent person.

3. When I am with my teacher, I feel loved and cared about.

4. When I am with my teacher, I often feel inadequate or incompetent.

5. When I am with my teacher, I have a say in what happens, and I can voice my opinion.

6. When I am with my teacher, I often feel a lot of distance in our relationship.

7. When I am with my teacher, I feel very capable and effective.

8. When I am with my teacher, I feel a lot of closeness and intimacy.

9. When I am with my teacher, I feel controlled and pressured to be certain ways.

Appendix I

Permission to use Classroom Engagement Inventory

RE: Classroom engagement inventory Inbox x

Wang, Ze 11:14 AM (4 minutes ago)
to me, Christi, David

Hi Dave,

Yes, you have my permission to use the CEI for research purpose. Attached is a copy of the CEI and the administration guide.

Good luck with your dissertation!

Ze Wang, Ph.D.
Associate Professor
Department of Educational, School, and Counseling Psychology
University of Missouri

Phone: (573) 882-7602
Email: WangZe@missouri.edu
Webpage: http://web.missouri.edu/~wangze

From: David Dennie [mailto:dennie.david@gmail.com]
Sent: Thursday, July 16, 2015 10:02 AM
To: Wang, Ze <WangZe@missouri.edu>
Cc: David Dennie <dennie.david@gmail.com>
Subject: Classroom engagement inventory

Dr. Wang, my name is Dave Dennie and I am a doctoral student at Columbus Statue University in Columbus Georgia. I am in the process of developing my dissertation proposal on teacher-student relationships, engagement, and student growth percentiles in Georgia and would like to request permission to utilize the Classroom Engagement Inventory in my dissertation. Can you please respond letting me know that is approved for me to use the CEI? Also, if you could, I would appreciate any supporting documentation that you have from Measuring Engagement in Fourth to Twelfth Grade Classrooms.

Thank you for any help you can provide and taking the time to read and respond to my email.

Dave Dennie

Appendix J

Permission to reprint Self-Systems Process Model

Re: Permission to use image School/Dissertation x

James Connell <james.connell@irre.org> Aug 1

to me

David: Feel free to use the image with attribution to the publication. Good luck with the dissertation and please let me know when it's available.

Best,

Jim Connell

On Sat, Aug 1, 2015 at 9:21 AM, David Dennie <dennie.david@gmail.com> wrote:

> Dr. Connell, my name is Dave Dennie and I am a doctoral student at Columbus State University in Columbus Georgia. I am in the process of developing my dissertation proposal on teacher-student relationships, basic psychological needs, engagement, and student growth percentiles in Georgia and would like to formally request permission to utilize the attached image which is located on pg. 51 from Competence autonomy and relatedness: A motivational analysis of self-system processes(1991) in my dissertation.
>
> Thank you for any help you can provide and taking the time to read and respond to my email. I look forward to hearing from you.
>
> Dave Dennie
> Columbus State University
> Doctoral Candidate

--

James P. Connell, Ph.D.

President

360-367-7710 phone